NASA Contractor Report 3610

Laboratory Studies of Scales for Measuring Helicopter Noise

J. B. Ollerhead

CONTRACT NAS1-16276
NOVEMBER 1982

NASA

NASA Contractor Report 3610

Laboratory Studies of Scales for Measuring Helicopter Noise

J. B. Ollerhead
Wyle Laboratories
El Segundo, California

Prepared for
Langley Research Center
under Contract NAS1-16276

NASA
National Aeronautics
and Space Administration

**Scientific and Technical
Information Branch**

1982

FOREWORD

It is not possible to mention all of the many people who contributed to this study, but the author would like to express particular appreciation to Mr. A. E. Clarke, United Kingdom Department of Industry, Dr. C. A. Powell, NASA Langley Research Center, Mr. T. G. Hargest, National Gas Turbine Establishment, Mr. L. C. Sutherland, Wyle Laboratories, Mr. P. Bradshaw and Mr. E. D. Rodgers, Loughborough University, Dr. J. W. Leverton, Westland Helicopters, and the members of ICAO CAN Working Group B. The experimental work performed at Loughborough and much preparatory work was largely supported by a grant from the Department of Industry.

SUMMARY

A laboratory study has been made of the adequacy of the Effective Perceived Noise Level (EPNL) procedure for rating helicopter noise annoyance. Recordings of 89 helicopters and 30 fixed-wing aircraft (CTOL) flyover sounds were rated with respect to annoyance by groups of approximately 40 subjects. The average annoyance scores were transformed to Annoyance Levels defined as the equally annoying sound levels of a fixed reference sound. The main experiment was performed at Loughborough University of Technology, England, using headphone presentation but a large part of it was repeated in the test facilities at Langley Research Center using loudspeaker presentation. The sound levels of the test sounds were measured on various scales, with and without corrections for duration, tones, and impulsiveness. On average, the helicopter sounds were judged equally annoying to CTOL sounds when their duration-corrected levels are approximately 2 dB higher. However, all scales predict the annoyance levels of helicopter noise significantly less consistently than those of CTOL noise, a finding which may be attributed to the widely differing acoustical characteristics of different helicopter types. Multiple regression analysis indicated that, provided the helicopter/CTOL difference of about 2 dB is taken into account, the particular linear combination of level, duration, and tone corrections inherent in EPNL is close to optimum. The results revealed no general requirement for special EPNL correction terms to penalize helicopter sounds which are particularly impulsive; apparently, impulsiveness causes spectral and temporal changes which themselves adequately amplify conventionally measured sound levels.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES (Continued)

1.0 INTRODUCTION

Aircraft noise certification standards have been specified by the FAA and the International Civil Aviation Organization (ICAO) for subsonic jet aircraft and for both large and small propeller-driven aircraft.[1,2] For the first two categories, noise limits are defined as Effective Perceived Noise Levels in EPNdB; for the latter, they are defined as Maximum A-Weighted Sound Levels, L_A, in dB(A).

In its deliberations to develop noise certification standards for V/STOL (vertical or short takeoff) aircraft including helicopters, Working Group B (WGB) of the ICAO Committee on Aircraft Noise (CAN) was concerned about evidence that these noise scales may be less satisfactory for rating helicopter noise than that of conventional takeoff and landing aircraft (CTOL).[3] Much of this evidence pointed to the possibility that in the case of helicopters, the existing noise scales might not properly account for the periodic impulsiveness which characterizes the sound of rotors. It is certainly widely acknowledged that severe forms of impulsiveness, often known as "blade slap," can be particularly intrusive and annoying, and it is clearly necessary that any noise scale used for certification should properly reflect the potential of such noise components to evoke annoyance.

The history of research into suitable helicopter noise rating methods is documented elsewhere (e.g., References 4 through 7). It suffices to state here that the evidence is contradictory; some studies have suggested that standard procedures such as $EPNL_t$* and L_A are adequate for V/STOL and helicopter noise while others indicate that they underestimate its noisiness.

Of particular significance, WGB asked the International Organization for Standardization (ISO) to study the problem of helicopter noise and recommend a suitable noise scale.[3] This work, some of which is described in Reference 8, was performed by Working Group 2 of ISO Technical Committee 43, Subcommittee 1, and culminated in the preparation of a draft ISO standard for helicopter noise measurement.** The main feature of this proposal was the adoption of a version of $EPNL_t$ modifed by a correction for impulsiveness (following the philosophy of the

*In this report, the standard version of Effective Perceived Noise Level which incorporates tone corrections is abbreviated $EPNL_t$ to distinguish it from an alternative version EPNL which does not.

**International Organization for Standardization (ISO), Draft Addendum ISO 3891/DADI, "Measurement of Noise from Helicopters for Certification Purposes," 1979.

"tone correction," another $EPNL_t$ modifier). The ISO impulsiveness descriptor is sensitive to large periodically occurring peaks in flyover sound pressure time history and augments $EPNL_t$ by up to 6 dB.

This descriptor was subsequently tested in a field experiment[9] at NASA's Wallops Flight Center in which two different helicopters and a propeller-driven CTOL aircraft were flown over a group of test subjects who compared their relative noisiness. When compared on the basis of $EPNL_t$ (without the ISO impulse correction), the two helicopters, a Bell 204B and a Bell OH58A, were judged equally noisy despite the fact that the 204B has a considerably more impulsive noise signature. This finding was broadly confirmed in laboratory experiments involving sound recordings made during the field trials.[10] In the light of this evidence, WGB concluded that the need for an impulse correction remained unproven and both ICAO and FAA consequently framed proposed helicopter noise certification procedures around the conventional $EPNL_t$ scale.[11, 12] The committee did, however, recognize a need for further research into the matter.

The present study was initiated during the period of deliberation in a further attempt to check the adequacy of $EPNL_t$ for the practical purposes of controlling helicopter noise. The main objective was to test and compare the abilities of a number of conventional noise rating scales to predict the relative annoyance levels of a wide range of recorded helicopter sounds and to identify components and characteristics of helicopter noise which contribute to annoyance but which may not be fully accounted for in the $EPNL_t$ model. Of special interest were (a) the relationships between helicopters and CTOL noise, (b) impulsiveness, and (c) the very long durations sometimes associated with helicopter flyover noise, particularly during the approach phase.

It is, of course, highly probable that many factors contribute to helicopter noise annoyance including both the acoustic qualities of the sound and nonacoustic information which the sound conveys. The precise role of each factor could only be established through extensive experiments in which each factor is varied independently of the others, either one at a time or simultaneously. The main requirements would be the correct identification and inclusion of all relevant independent variables and, as the name implies, independence of these variables.

Theoretically, single factors such as impulsiveness can be studied through relatively small scale experiments in which this factor is the only physical variable. In practice, it is often difficult, if not impossible, to vary a single factor independently of all others. For example, a change of impulsivity normally causes a change in the frequency spectrum. In the case of helicopter noise, impulsivity may also be associated with increased duration, as will be seen. This "confounding" of factors is difficult to unravel and the isolation of a satisfactory noise rating scale may only be possible through a trial-and-error process in which the model is evaluated and refined by testing it against new experimental data as they become available.

The basic approach to this study was to gather together a large collection of helicopter noise recordings from which a test sample could be selected to cover wide but realistic variations of at least the major variables of interest (duration, tonality, and impulsiveness). Each sound would be rated with respect to its annoyance-evoking qualities by a group of test subjects and measured on various standard scales of noise measurement including A-weighted sound level (L_A) and Effective Perceived Noise Level ($EPNL_t$). The performance of these scales as annoyance predictors could then be assessed by comparing the measured sound levels and the subjective "annoyance levels." If a sufficiently large and varied sample of sounds were available, then it would also be theoretically possible to isolate directly the independent contributions of these variables to judged annoyance by appropriate multivariate statistical methods.

Certain difficulties associated with this kind of experimentation were recognized at the outset. Foremost among them is that reliance upon available recordings of real aircraft flyover sounds imposes severe constraints upon the variations of, and relationships between, variables of importance. It might be possible to achieve a reasonable degree of decorrelation between a few primary variables but many subsidiary variables including variations of the signal with time, Doppler frequency shifts, rotor blade passing frequencies, and many others which may affect a listener's assessment of a particular event, inevitably lie beyond the control of the experimenter. As noted previously, elaborate annoyance prediction models to account for many such factors could only be synthesized on the basis of results from highly controlled experiments in which those factors are varied systematically.

Indeed, it was on systematic experiments of this kind that the foundations of $EPNL_t$ were laid and from which emerged duration and tone corrections and more recently the ISO impulsivity correction. However, it is by no means clear that this process is entirely satisfactory when conducted in isolation. A fairly extensive test of $EPNL_t$ made by the author[13] revealed certain deficiencies which, although of little consequence when the scale is used to compare aircraft of similar performance and acoustical characteristics, suggested that it would be unwise to place too much reliance on $EPNL_t$ for the purposes of comparing the perceived noisiness of very dissimilar aircraft. The results pointed to the need for the more systematic experiments to be accompanied by practical evaluation of psychoacoustical models through tests such as those described here.

The original program plan called for the inclusion of up to 200 individual helicopter flyover recordings. These were to be evaluated in subjective tests at Loughborough using headphone presentation and subsequently at Langley Research Center using loudspeaker presentation. This very large sample of test sounds was considered practicable through the use of a fast rating scale test procedure to obtain annoyance assessments of each sound.

Because less than 200 original sound recordings were obtained and because of other difficulties, the scope of the experiments had to be curtailed. In an attempt to compensate for this to some extent, a large part of the basic experiments was duplicated in three independent tests; one at Loughborough, again using headphone presentation, and the others in two separate test facilities at Langley Research Center using loudspeaker presentation.

The use of headphones offers numerous advantages over loudspeakers: closer control over variations in sound level and frequency response, comfortable and convenient surroundings for the test subjects, and the ability to handle large numbers of subjects at a time. The disadvantages include the difficulty of accurately measuring the test stimuli and uncertainties concerning the relationships between normal free field or diffuse listening conditions and the pressure field of the headphones. A check on present headphone results using loudspeakers was therefore felt to be desirable.

In this report, the main experiment is described in detail in Sections 2, 3, and 4. This is followed in Sections 5 and 6 by a description of the duplicate experiments and the overall conclusions. Appendices contain (A) the Instructions to the Subjects, (B) a summary of the acoustic characteristics of the test sounds, (C) representative time histories and spectrum plots of some of the helicopter test sounds, and (D) a summary of basic characteristics of most of the helicopters utilized for recordings employed in this program.

2.0 PRELIMINARY STUDIES

Before carrying out the main experiment, a number of preliminary tests were conducted to investigate certain aspects of the proposed test procedures. These involved headphone presentation of recorded test sounds to groups of between 36 and 40 subjects.

2.1 The Subjective Rating Scale

The basic laboratory method used to measure subjective reactions to helicopter noise employs a simple but efficient rating scale procedure used previously by Powell[10] and others. The test subjects are simply asked to listen to each sound in turn and to rate its annoying qualities on a continuous numerical rating scale ranging from 0, labeled (for example) "Not Annoying At All," to 10, labeled "Extremely Annoying." Example test instructions are given in Appendix A.

Limitations of simple rating scales of this kind are well documented. In particular, they are normally considered unlikely to meet the requirements for a true interval scale making the applicability of parametric statistics somewhat dubious. Also, different subjects distribute their scores differently along the scale. Finally, the subject has no absolute point of reference so that scale values may be assigned differently in different tests.

The procedure was tested by comparing it with the Method of Adjustment, one of the oldest and most fundamental of the psychophysical testing methods and certainly one which has been well-tested in the field of auditory perception.

Twenty-two different aircraft flyover sound recordings were rated with respect to noisiness by the same 36 subjects using each of the two methods in two separate experiments. The sounds covered the range 65 to 90 dB(A) (approximately). In the adjustment experiment the subjects heard each test sound at a fixed level alternating repeatedly with a reference flyover sound whose level they could control. They were asked to adjust the level of the reference sound until the two sounds were considered equally noisy.* In the rating experiments, the subjects listened to the test sounds in a random sequence and assigned each one a noisiness value between 0 and 10 on their score sheets.

Figure 1 shows the average subjective (rating) scores SS plotted against the average adjusted levels AL. The straight line is the regression of SS upon AL and

*"Noisy" was defined as "unwanted, objectionable, disturbing or unpleasant."

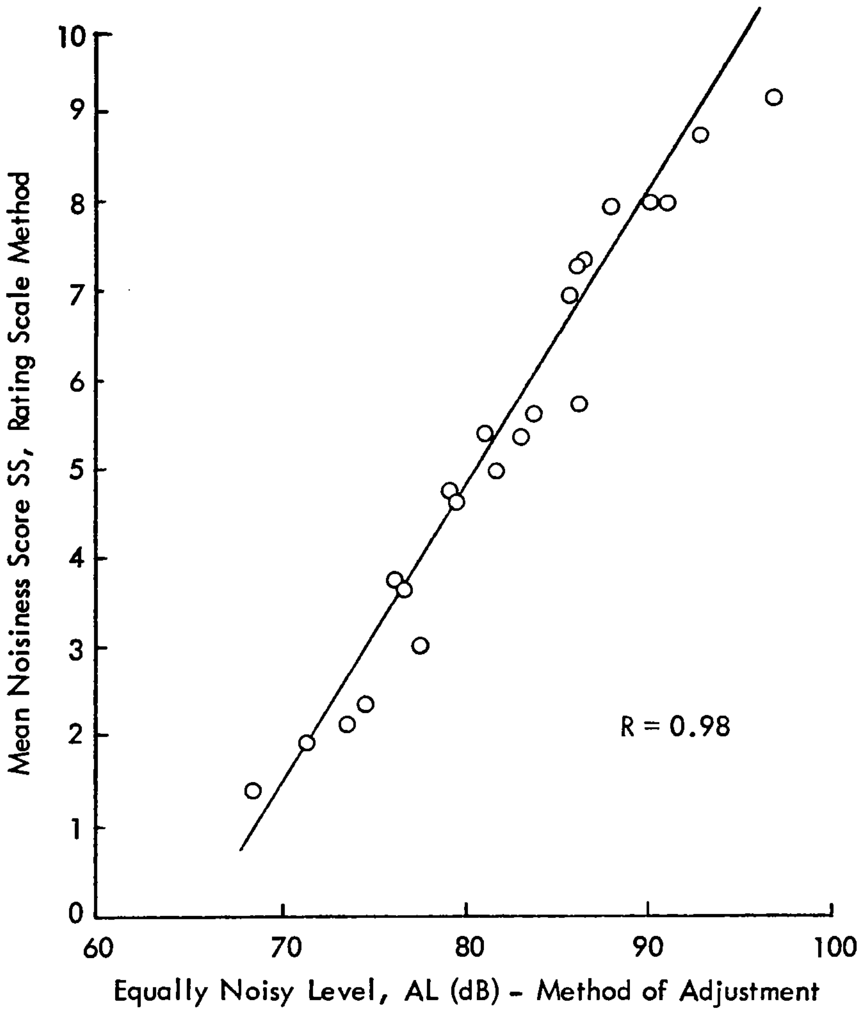


Figure 1. Correlation Between Rating Scale Method and Method of Adjustment

the standard deviation of the points about the line in the y-direction is 0.51 scale units. The correlation coefficient is 0.98. In fact, it is evident that the relationship is really slightly nonlinear with the subjects "running out of numbers" at the extremes of the rating scale and a nonlinear regression would doubtless provide a slightly better fit. Analysis of the variance of individual scores showed that the adjustment method yielded rather smaller standard errors of the mean scores (when expressed in relation to the total variance for all sounds). However, despite the potential drawbacks of the rating scale method, the high correlation between the two sets of mean results substantiates its reliability for present purposes where it offers the considerable advantage of being an order of magnitude faster to administer.

2.2 Calculation of Annoyance Levels

Although the use of the Rating Scale method allows rapid evaluation of a large number of test sounds, two limitations, suggested earlier, had to be considered. The first was that although scores from a single test session may be compared directly, the rating scale has no fixed point of reference to allow valid intertest comparisons. The second was that if the scale is not linear, the results may not be amenable to parametric statistical analysis.

These possible difficulties were circumvented by including in each test session a number of repetitions of the same single reference sound (itself an aircraft flyover) played at different levels. For each test, the response scale could thus be calibrated in terms of the sound level of the equally noisy reference sound, in dB.

A pilot experiment was performed to test this procedure. A test tape of approximately 30 minutes duration was played to 40 subjects through headphones. The tape contained 33 test sounds comprising a variety of helicopter flyover recordings of various levels and durations. The reference sound which was randomly intermixed at eight different sound levels was a 20-second long recording of a T-28 single piston-engined propeller aircraft flying overhead at a height of 90 m. (This T-28 was used in the NASA Wallops Island experiment;[10] this particular sound was also used as a reference in all subsequent tests.)

Figure 2 shows the relationship between the mean subjective scores SS and the nominal level of the reference sound in dB(A) (measured at input to the headphones). The relationship is clearly linear with a high correlation (and is

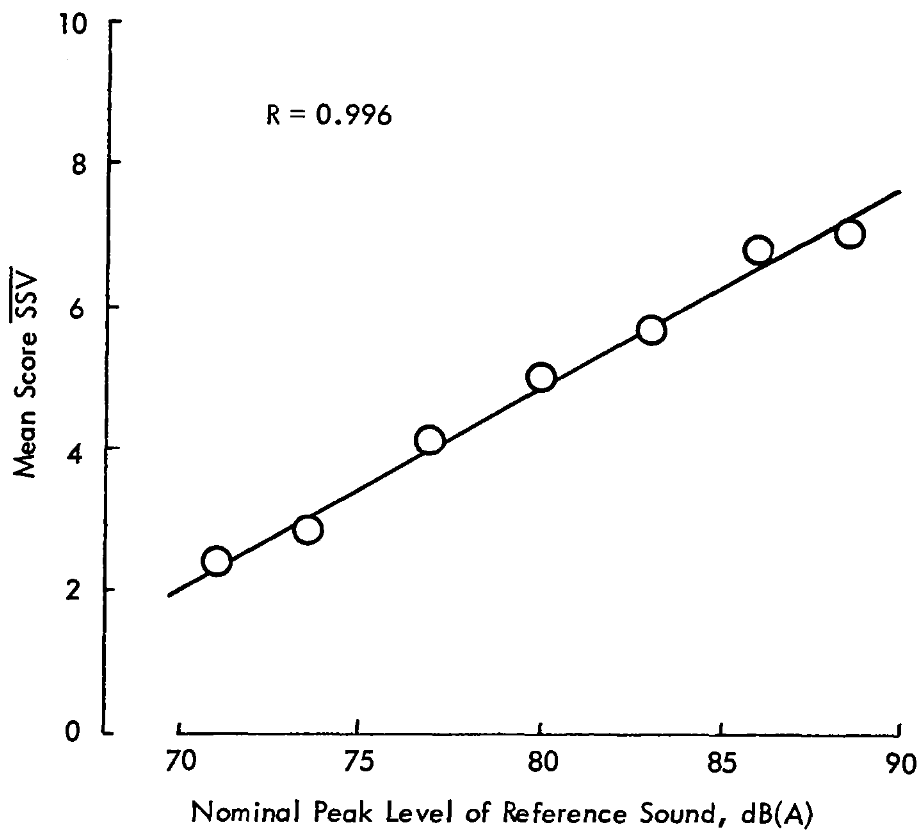


Figure 2. Relationship Between Mean Subjective Ratings and Nominal Peak Level of Reference Sound

typical of those found in all subsequent tests). The regression line was used to transform the average subjective score (SS) for each of the test sounds to a "Noisiness Level" NL in dB(A). On this scale, the standard error of the mean scores is approximately 1 dB. In subsequent experiments, which involved judgments of annoyance rather than noisiness, the transformation is termed "Annoyance Level."

2.3 Consistency of the Scaling Procedure

Although a balanced presentation order scheme is desirable in which different subjects or groups of subjects hear the test sounds in a different sequence, this was not practicable within the scope of this study. Instead, the single presentation sequence was randomized (including the reference sounds) and various tests were made to check the consistency of the results.

In the preliminary tests, these included (a) repetition of the first four sounds at the end of the test, and (b) repetition of one sound at three (random) points during the main part of the test. No significant differences were found between the mean scores for any of these repetitions. The fact that SS is highly correlated with the level of the reference sound (R = 0.996, see Figure 2) also indicates that the test method gives consistent results.

In the main experiment, the possibility of an order effect (due to any tendency for subjects when scoring to be influenced by their memories of the previous test sounds) was examined by computing the serial correlation between subjective scores (i.e., the correlation between the scores for all test sounds and those of their preceding sounds). No significant correlation was found.

2.4 Contribution of the Approach Noise Component to Annoyance

In order to probe the role of the long approach phase associated with some helicopter flyover sounds the pilot tests included the following three sounds in which the approach components were modified:

1. Bell 205 flyover at 120 kt and 150 m altitude, specially recorded for this study by NASA Langley Research Center (LaRC). This sound exhibited heavy blade slap during the approach and had a "10 dB-down" duration of 19 seconds. Two versions were included in the test: 1A - the entire recording with a total duration of 82 seconds, and 1B - the 10 dB-down flyover segment only (of 28 seconds total duration including ramps at the beginning and end).

2. OH58A flyover at 90 m altitude. This had an unintrusive main rotor component and a 10 dB-down duration of 9 seconds. Again, two versions of this sound were included: 2A - the complete sound of 41 seconds duration, and 2B - the flyover component only of 17 seconds duration.

3. Bell 205 flyover at an altitude of 150 m and a speed of 40 kt. This sound was also recorded specially and had a very long approach phase during which the thumping of the main rotor was particularly noticeable. The impulsiveness became very harsh during the final stages of the approach, developing into a cracking sound just before the overhead position. The 10 dB-down duration was 52 seconds. Three versions were used:

   3A. The complete sound covering a dynamic range of 30 dB; duration 207 seconds;

   3B. approximately the upper 20 dB of the signal; duration 144 seconds; and

   3C. approximately the upper 10 dB; duration 59 seconds. (This was presented three times during the main part of the test.)

The average subjective (noisiness) scores (and their standard deviations) for these seven sounds were as follows:

1A. 6.8 (1.7)
1B. 6.8 (1.6)

2A. 5.6 (1.8)
2B. 5.6 (1.8)

3A. 6.7 (1.8)
3B. 6.0 (1.8)
3C. 6.2 (1.5)/6.8 (1.7)/6.6 (1.9)

The differences between the mean scores for the different versions of each sound are not significant at the 5 percent level, indicating that the sound outside the 10 dB-down limits does not contribute significantly to perceived noisiness as measured in this experiment.

To investigate the possibility that the written instructions caused the subjects to focus their attention on the loudest part of the flyover sound (and not upon the approach component), the same test was repeated with a different group of 36 subjects and with slightly modified instructions. In this case, the subjects were specifically asked to:

> "... consider how you would feel if you heard (the sound) at home on a number of occasions during the day and take account of all the characteristics of the sound including its duration."

Furthermore, in a second repeat test with a third group of test subjects, the same test tape was used but the instructions were based on "annoyance" rather than "noisiness" to explore the possibility that subjects' ratings of annoyance were less influenced by loudness than were noisiness judgments. No duration cues were included. The results of these two further pilot experiments agreed closely with those of the first. When the subjective scores were transformed to noisiness or annoyance levels (Figure 3 compares the transformation curves for the three tests), only in the cases of two sounds were inter-test differences significant ($p = 0.05$) and these were the second and fifth sounds of the test. These results did not alter the conclusion that the rated noisiness or annoyance levels were not influenced by the approach component.

2.5 Conclusions

On the basis of these preliminary tests, it was concluded that:

i. Noisiness and annoyance may be considered as interchangeable attributes for present purposes; annoyance judgments are slightly preferable because the instructions are simpler.

ii. The 11-point rating scale bears a highly linear relationship to the sound level of an equally noisy or annoying standard reference sound. A poor range of reference sound levels might cause some skewing of this relationship but transformation to NL should still provide a cardinal annoyance scale and allow combination of results from different tests.

iii. With 40 subjects, the average standard error of mean judgments is equivalent to 1 dB (NL) and repeatability is good, both within a single test and between tests.

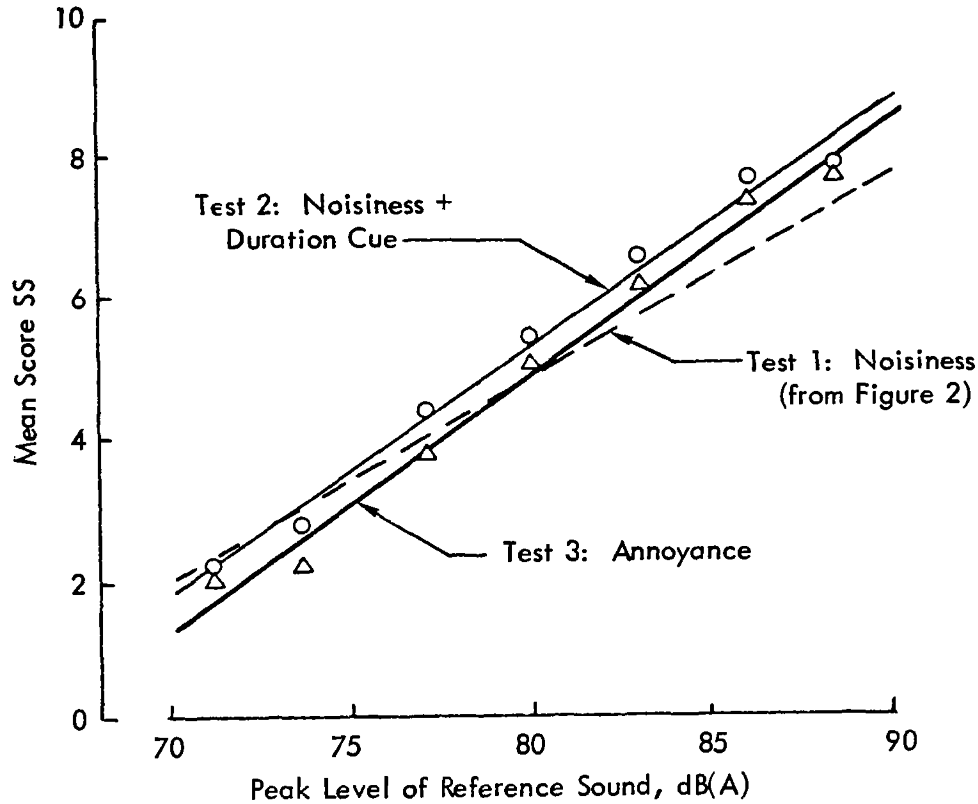


Figure 3. Comparison of "Calibration Curves" from Three Preliminary Tests

iv. The basic test procedure is stable and insensitive to variations in the subjects' instructions.

v. Even in a very extreme case (the 3.5 minute long recording of the Bell 205 approach), the "approach component" (before the first 10 dB-down point) makes no measurable contribution to judged noisiness or annoyance. It is therefore unlikely that any influence of the approach component will be isolated through an experiment which does not take account of background noise masking effects. Accordingly, it was decided not to include approach component duration as a variable in the main tests.

3.0 DESCRIPTION OF THE MAIN EXPERIMENT

3.1 Test Tapes

The main tests involved an evaluation of 119 aircraft sounds; 89 helicopters and 30 CTOLs* which are described in Appendix B. The helicopter recordings were selected from approximately 140 available to provide the widest possible range of types and flight conditions as well as satisfying the requirements of reproduction quality. See Appendix C for representative time histories and spectra and Appendix D for general characteristics for the helicopters included.

Most of the helicopter flights were level flyovers although some recordings were made during approach descents. The CTOL's, which were included to allow direct comparison of the relative performance of the noise rating scales as applied to helicopters and fixed-wing aircraft, were recorded for this study at London (Heathrow) Airport at positions close to the nominal approach and flyover certification points.

The sounds were rerecorded in random sequence onto four test tapes. Each tape of approximately 30 minutes duration contained a total of 44 flyover sounds including eight reference sounds (the same T-28 flyover recording used in the preliminary experiments) recorded at 3 dB intervals over a dynamic range of 21 dB and the same five sounds recorded at the beginning and end of the tape (results for the first five were discarded to minimize the effects of any initial period of adjustment or adaptation by the subjects).

The test sounds were recorded on, and replayed from, a Nagra IV S tape recorder running at 7-1/2 ips. All sounds were manually "ramped" at start and finish and the interval between sounds was about 8 seconds during which a voice announcement of the next sound number was recorded (although in most test runs this was suppressed in favor of an electronically-controlled digital display).

The test tape was replayed to six subjects at a time through Koss PRO 4AA headphones driven by six specially constructed power amplifiers. A control unit mixed the test signals with a very low level broadband background sound whose function was to mask perceptible switching transients between sounds. The same unit suppressed the voice announcements and operated individual sound number

*Conventional takeoff and landing aircraft - in this case, all transport category types, mostly turbofan-powered.

displays when these were in use. This process was controlled by a 12 kHz tone recorded on the second tape recorder channel. To eliminate slight cross-talk during replay, the data channel was low-pass filtered at 8 kHz.

3.2 Test Procedures

The four test tapes were administered to between 36 and 40 test subjects, most of whom were undergraduate students in the age range 19 to 23 with roughly equal numbers of males and females.

The test subjects sat in armchairs inside a quiet test room. Written instructions read by the subjects before a test together with a score sheet are presented in Appendix A. The instructions were verbally reinforced and the broad purpose of the test was also explained. Most subjects participated in three tests on three separate occasions but prior to the first they were given a practice test comprising six typical sounds covering the sound level range to be heard subsequently. Subjects recorded their scores for each sound by marking numbers on their score sheets between 0 and 10. In most tests, the sound number was continuously presented on small LCD display units affixed to their clipboards.

3.3 Noise Levels

The sound recordings were analyzed to yield measurements on the various scales of noise level summarized in Table I, taking account of the frequency response of the headphones. Real-time one-third octave band analysis was performed on a GenRad 1921 analyzer coupled to a PDP 11/34 computer. The data reduction program incorporated a frequency response correction function which was based on the average response for the 12 individual earphones used in the tests.

To obtain this function, individual headphone output levels were measured underneath the headphone cushion on the head using a Knowles miniature microphone and a "pink noise" input. The frequency response of the miniature microphone was in turn measured by calibrating it against a pressure response condenser microphone in a flat-plate headphone coupler. The average frequency response together with the standard deviation for the 12 headphones is shown in Figure 4.

Because it was not possible to measure accurately the sound pressure inside the headphones when in normal use, the impulse correction terms were computed from the A-weighted tape recorder output. An indication of the likely effect of

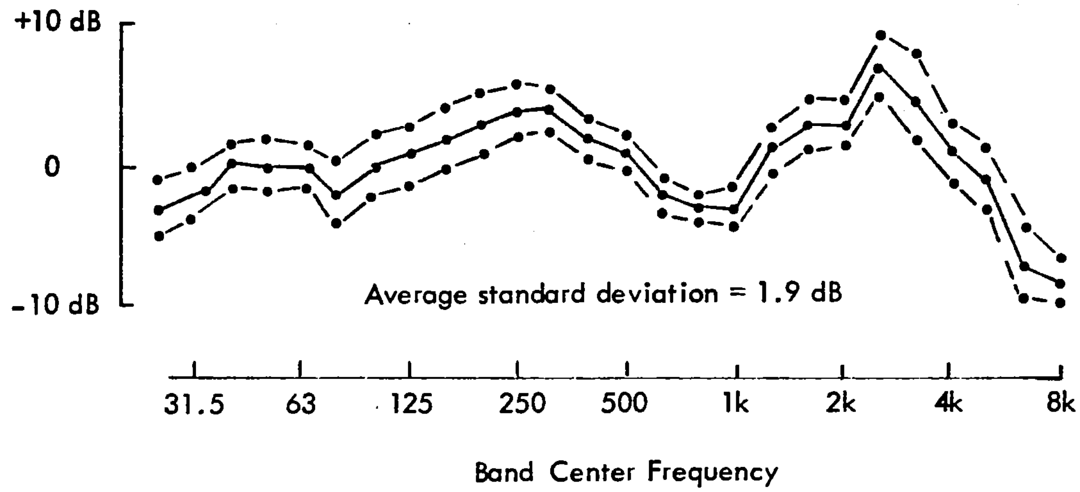


Figure 4. Mean Frequency Response of Headphones (12 Ears) Showing ±1 Standard Deviation

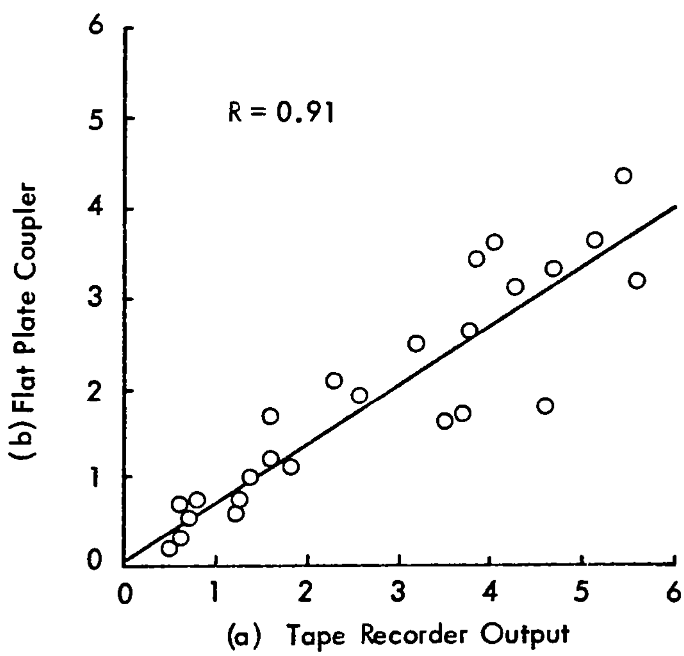


Figure 5. Comparison of Impulse Corrections ($EPNL_{ti}$ - $EPNL_t$) As Measured from (a) Tape Recorder Output and (b) in Flat Plate Headphone Coupler

the headphones upon impulsiveness can be gained from a sample of 25 measurements made with a pressure response microphone in the flat plate coupler. Figure 5 compares the impulse correction terms ($EPNL_{ti}$ - $EPNL_t$) as computed with the 1/2-second values measured in the two different ways. On average, the coupler values are approximately two-thirds of the direct values with a standard deviation of 0.3 dB. Repetitions of the headphone measurements for one particular flyover recording (S61N1) using 10 different headphones showed very little variance in the magnitude of the average impulse correction (standard deviation = 0.1 dB).

For each basic scale, two levels were computed: (a) the maximum 1/2-second value during the event, and (b) the time-integrated or "duration corrected" value obtained by the summation process incorporated into the $EPNL_t$ procedure which covers the upper 10 dB of the time history.[1, 2] Time-integrated (i.e., duration corrected) levels are denoted by abbreviations prefixed by the letter "E". It should be noted that the weighted sound pressure levels were computed from the one-third octave band level arrays using the weighting functions listed in Table 2 and plotted in Figure 6.

Table 1

Noise Level Scales

Abbreviation	Description
L (EL)	Overall sound pressure level, dB
L_A (EL_A)	A-weighted sound level, dB(A)
L_D (EL_D)	D-weighted sound level, dB(D)
L_E (EL_E)	E-weighted sound level, dB(E)
L_F (EL_F)	"F"-weighted sound level, dB(F)
PNL (EPNL)	Perceived Noise Level, excluding tone correction, PNdB
PNL_t ($EPNL_t$)	Perceived Noise Level, including tone correction ($EPNL_t$ is the Standard ICAO version)
PNL_{ti} ($EPNL_{ti}$)	PNL with tone correction and ISO impulse correction
PNL_{tc} ($EPNL_{tc}$)	PNL with tone and crest factor impulse correction
LA_c (EL_{Ac})	A-weighted sound level with crest factor impulse correction

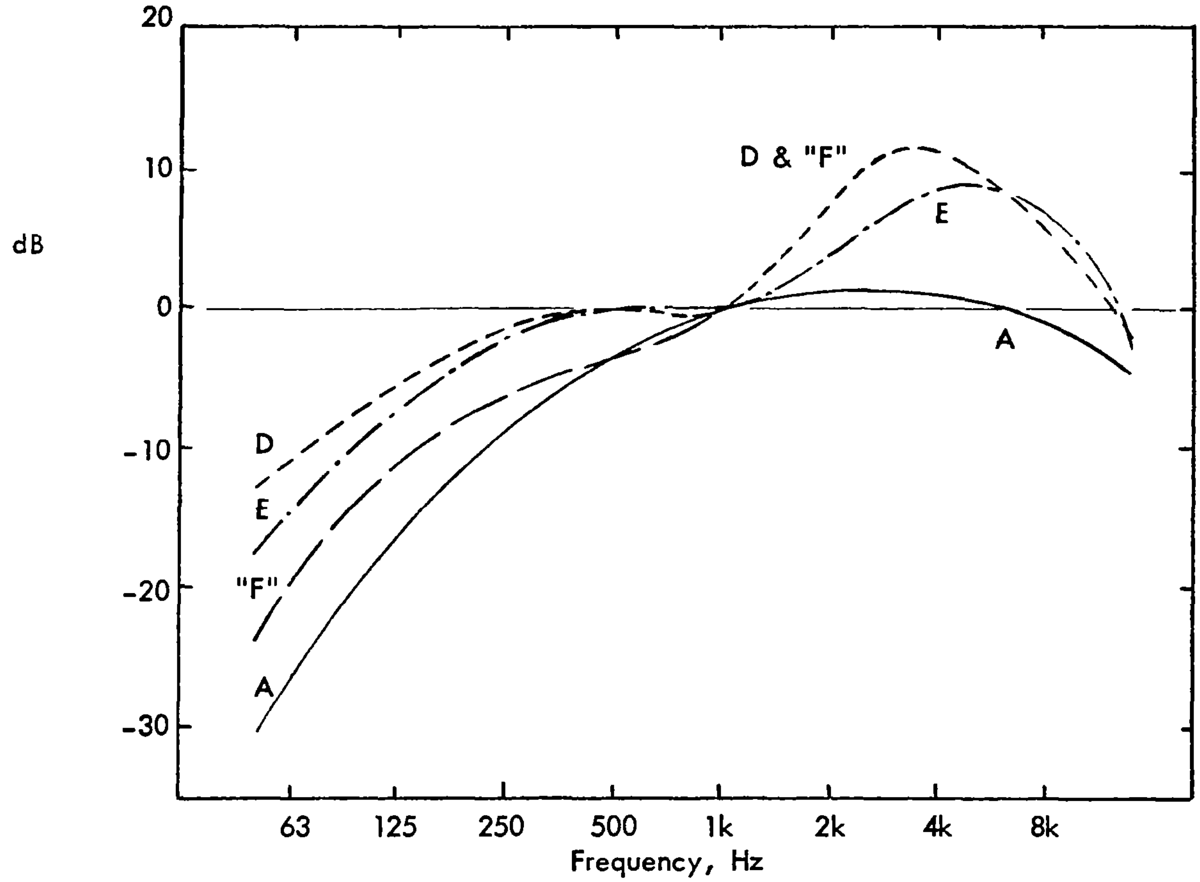


Figure 6. Sound Level Frequency Weighting Curves

Table 2

Sound Level Weighting Functions

Frequency, Hz	Weighting, dB			
	A	D	E	"F"
50	-30.2	-12.8	-17.4	-23.8
63	-26.2	-10.9	-14.5	-19.5
80	-22.5	-9.0	-11.8	-15.9
100	-19.1	-7.2	-9.4	-13.1
125	-16.1	-5.5	-7.3	-10.9
160	-13.4	-4.0	-5.3	-8.8
200	-10.9	-2.6	-3.6	-7.4
250	-8.6	-1.6	-2.2	-6.2
315	-6.6	-0.8	-1.1	-5.2
400	-4.8	-0.4	-0.3	-4.3
500	-3.2	-0.3	0.1	-3.4
630	-1.9	-0.5	0.1	-2.5
800	-0.8	-0.6	0	-1.4
1,000	0	0	0	0
1,250	+0.6	2.0	0.7	2.0
1,600	+1.0	4.9	2.1	4.9
2,000	+1.2	7.9	4.0	7.9
2,500	+1.3	10.6	5.9	10.6
3,150	+1.2	11.5	7.6	11.5
4,000	+1.0	11.1	8.7	11.1
5,000	+0.5	9.6	9.1	9.6
6,300	-0.1	7.6	8.3	7.6
8,000	-1.1	5.5	6.5	5.5
10,000	-2.5	3.4	3.8	3.4

The E-weighted scale is based on Steven's generalized "perceived level" function.[14] The F-weighting is the (nonstandard) abbreviation assigned to a curve derived by Powell from a study of the relations between impulsiveness, repetition rate, and judged annoyance of simulated helicopter sounds.[15] Above 1 kHz, it is identical to the D-weighting. Below 1 kHz it rolls off more rapidly than the D-curve approaching the A-curve at the very lowest frequencies (see Figure 6 and Table 2).

The ISO impulsiveness correction* is applied to the half-second sound level time history of the flyover sound in a manner analogous to the use of the tone correction. The correction is computed from the A-weighted sound pressure time history p(t) which is low-pass filtered at 2 kHz for anti-aliasing purposes and digitized at 5 kHz. For each half-second time period, a quantity X is computed where

$$X = 10 \log_{10} \left[\frac{1}{N} \sum_{i=1}^{N} \left(\frac{p_i^2 - s}{s} \right)^2 \right]$$

$$s = \frac{1}{N} \sum_{i=1}^{N} p_i^2$$

and p_i are the N sampled values of p(t). The half-second impulsiveness correction is then given as follows:

if $X < 5.5$	$\Delta = 0$
if $5.5 \leq X \leq$ to 10.5	$\Delta = 0.8\,(X-3)$, dB
if $10.5 < X$	$\Delta = 6$ dB.

The "crest factor impulse correction" is also computed from the digitized A-weighted sound pressure time history. A crest factor C is calculated for each half-second period as the ratio

$$C = \frac{p_{max}^2}{\frac{1}{N} \sum p_i^2}$$

where p_{max} is the largest numerical value of p_i. The impulse corrected level is then given by (for PNL_{tc} for example):

*International Organization for Standardization (ISO), Draft Addendum ISO 3891/DADI, "Measurement of Noise from Helicopters for Certification Purposes," 1979.

$$PNL_{tc} = PNL_t + 10 \log_{10} C - 12$$

subject to the proviso that $PNL_{tc} \geq PNL_t$ (subtraction of 12 ensures that PNL_{tc} = PNL_t for broadband random noise).

It must be pointed out that all noise level calculations can only be considered approximate in that (a) the weighted levels are computed from one-third octave band levels, (b) although the time integration periods are nominally 0.5 second, they were in practice controlled by the cycle time of the GR 1921 analyzer which is slightly less than this (a difference which is, of course, accounted for in the integration process), and (c) the impulse correction is also nominal rather than actual because it does not allow for unmeasurable differences caused by the headphone response. Although these approximations mean that all calculations strictly are "nonstandard," the effects of (a) and (b) are considered to be negligibly small. The magnitude of the error due to (c) which is significant cannot be estimated with any precision although we may be confident that in general the true impulsiveness will be somewhat less than the nominal value.

3.4 Annoyance Levels

The mean subjective score SS (and standard deviation) for each sound were calculated across all subjects. For each test, the value of SS was plotted against measured levels L_A and $EPNL_t$ for the eight repetitions of the reference sound and the regression lines were then used to convert SS for each test sound to Annoyance Levels NL, in dB(A), and NLE, in EPNdB. In other words, the Annoyance Levels, NL and NLE, of any sound are the levels (in dB(A) and EPNdB) of the standard reference sound which would be equally annoying. NLE was included to make suitable allowance for possible nonlinearity between dB(A) and EPNdB over the wide dynamic range of the tests. (In fact, the relationship was entirely linear for the reference sound with the relationship NLE = NL + 9.0.)

3.5 Accuracy Considerations

The accuracy of the experimental method can be assessed in two ways. The square root of the grand average inter-subject variance (averaged across all test sounds) is 1.5 annoyance scale units which yields a standard error of 0.25.* Since one annoyance unit translates to approximately 4 dB on the NL scale, this may be interpreted as an average standard error of approximately 1 dB, i.e., the 95 percent confidence interval associated with any individual NL is about ± 2 dB.

A check on this is provided by the annoyance scores for the standard reference sound which is repeated through the main part of the tests 32 times (albeit at different levels). The average standard error of estimate about the regression lines** may therefore be taken as a measure of the variability of individual NL values. This has a value of 1.4 dB. This is a little larger than the standard error computed above but the difference could be explained by the small sample size.*** These considerations suggest that errors (i.e., the standard deviation) associated with a perfect noise rating scale would not be less than about 1.5 dB in this experiment.

*For 36 subjects.

**The lines used to convert from SS to NL.

***A standard F-test shows that there is about a 1 percent probability that this difference arose by chance.

4.0 RESULTS OF THE MAIN EXPERIMENT

4.1 Analytical Considerations

Sample results are shown in Figures 7 and 8 in the form of "scatter diagrams" of measured level plotted against annoyance level (the significance of the different plotting symbols will be discussed later). The correlation between measured (y) and judged (x) levels may be expressed in various ways and a choice depends upon the criteria of assessment, especially concerning the linearity of underlying relationships. It might be supposed for example that since both ordinate and abscissa in Figure 7(a) are maximum levels, expressed in dB(A), the underlying relationship should be the line $y = x$. Figure 7(a) shows that this is clearly not the case.

This discrepancy suggests that L_A is not a particularly good estimator of NL for the test sounds in general. But the form and magnitude of the apparent error depends on the precise choice of reference sound (which itself should be assigned no more importance than any one of the individual sounds in Figure 7(a)) and the gross deviation of the data cluster from $y = x$ may depend upon special peculiarities of the one used. On the other hand, the many test sounds may vary with respect to factors of importance not accounted for in the variable L_A.

The ultimate purpose of the noise measurement scales under investigation is to predict average annoyance levels. In this respect, it would be more logical to reverse the axes in scatter diagrams like Figure 7(a). However, in these tests of the predictive performance of the scales, annoyance level NL is the independent variable (admittedly involving a degree of experimental error) and the dispersion of the data points in the y-direction is a measure of how well (or how poorly) the noise measurement scales do their job.

Of course, any noise measurement scale for which the data points are clustered tightly about a monotonically increasing relationship between y and x may be considered good for the practical purposes of rating aircraft noise. However, in the context of the present tests, it is also considered desirable that the relationship between measured level and annoyance level be constrained to be linear with unit slope. This is because the only property of the reference sound which changes significantly with NL or NLE is that of sound level itself. Any composite noise scale which purports to take proper account of temporal and spectral variations in the test sounds should by definition incorporate the appropriate tradeoffs between their contributions and that of sound level, maintaining the relationship $y = x + c$. The constant c should ideally be zero or at least small.

The best fitting straight line of unit slope (i.e., the unit slope line about which the variance in the y-direction is minimized) passes through the centroid of the scatter diagram so that the constant c is the mean value of the error $(y_i - x_i)$. The goodness of fit is inversely related to the standard deviation s of the error. In Figure 7(a) the unit slope line is $y = x - 4.6$ with a standard deviation $s = 2.5$ dB.

4.2 General Comparison of Noise Level Scales as Predictors of Annoyance Level

Scatter diagrams comprising plots of measured level against annoyance level for some of the various noise measurement scales are presented as Figures 7(a) through 8(g). Different plotting symbols are used for the subgroupings identified in Table 3.

Table 3

Subgroups of Data in Main Experiment

Subgroup	Sample Size
Less Impulsive Helicopters	73
More Impulsive Helicopters	16
CTOL Approaches	12
CTOL Takeoffs	18

The "more impulsive" helicopter sounds are those for which the integrated impulse correction given by $I = EPNL_{ti} - EPNL_t$ is greater than or equal to an arbitrary threshold value of 4 dB.

The unit slope straight line in each diagram is fitted to all 119 data points to minimize the error variance in the y-direction.* Table 4 lists the overall mean prediction error and its standard deviation together with the mean and standard deviation of the displacements of each data subgroup from the overall mean line. Thus, for example, the mean error L_A - NL for all 119 sounds is -4.6 dB with a standard deviation of 2.5 dB. The 89 helicopter points lie on average 0.2 dB below this mean error line (standard deviation = 2.6 dB) and the 30 CTOL points lie on average 0.3 dB above it (standard deviation = 1.9 dB). The further breakdowns in Table 4 give the margins for "more" and "less" impulsive helicopters separately and

*i.e., the line $y = n + c$ is positioned so as to minimize the dispersion of the data points about it in the vertical (y) direction. This dispersion will be greater than that about the linear regression line (of y on x) if the slope of the latter is not unity.

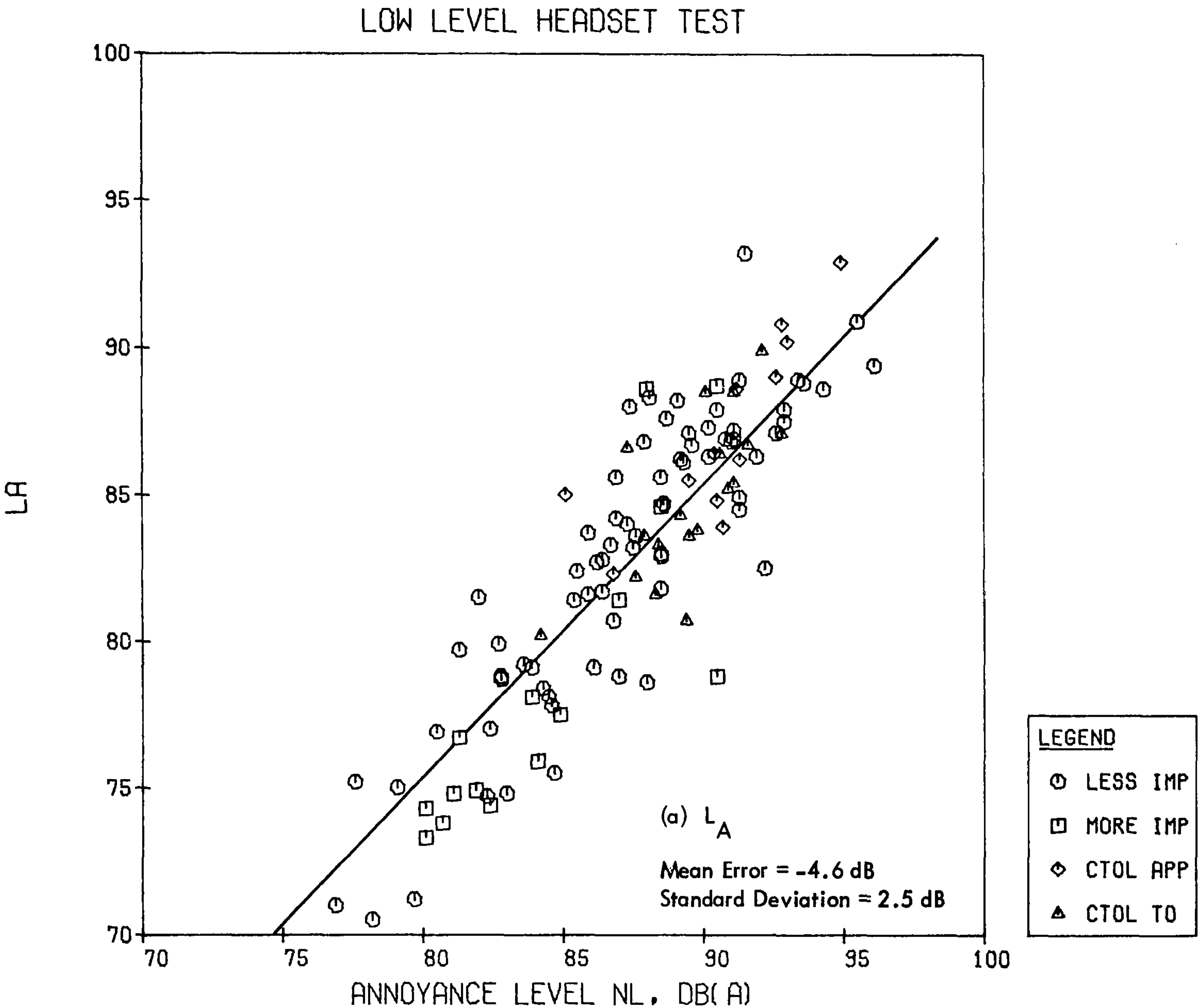


Figure 7. Measured Levels Versus Judged Annoyance Levels; Maximum Level Scales

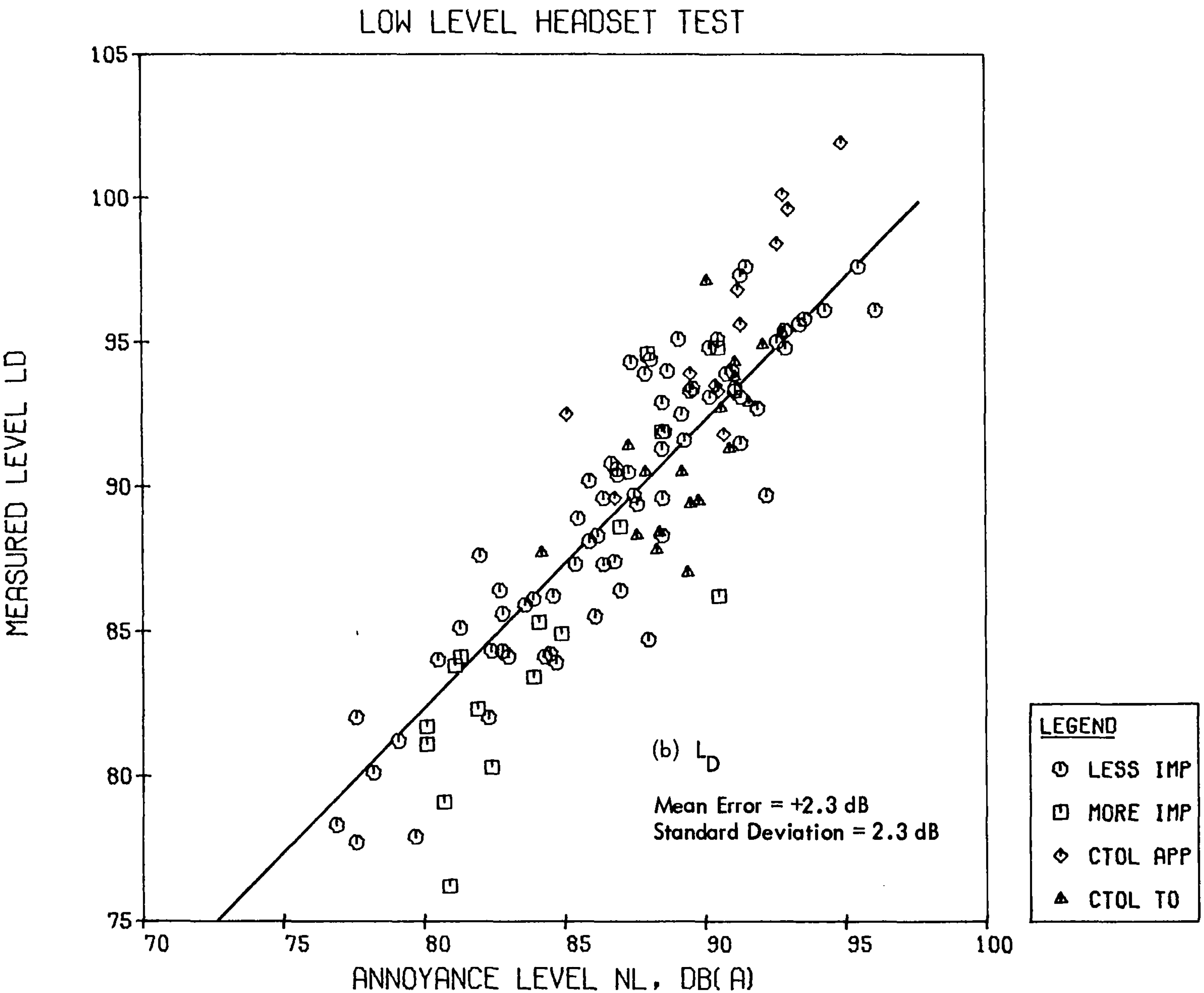


Figure 7. (Continued)

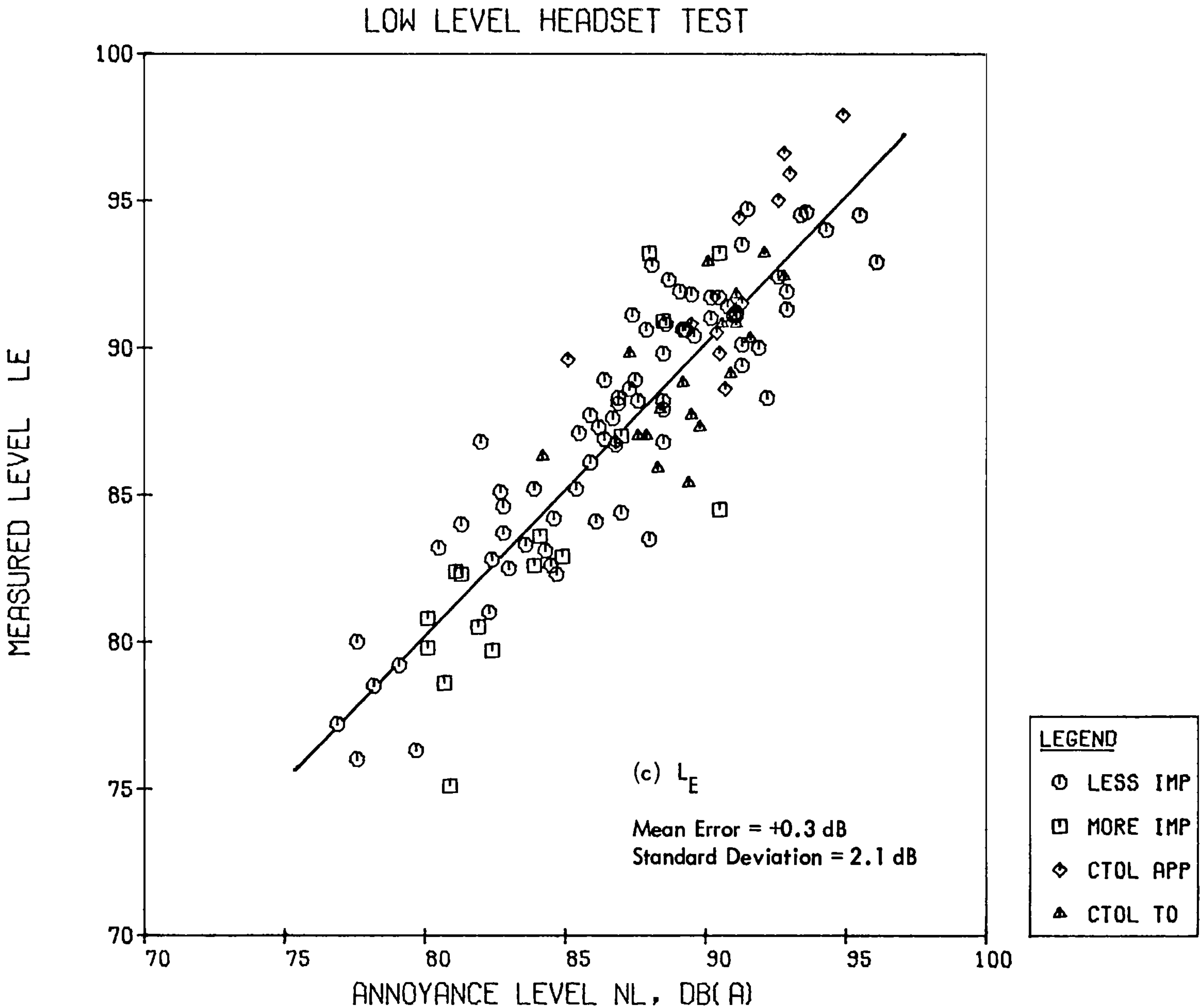


Figure 7. (Continued)

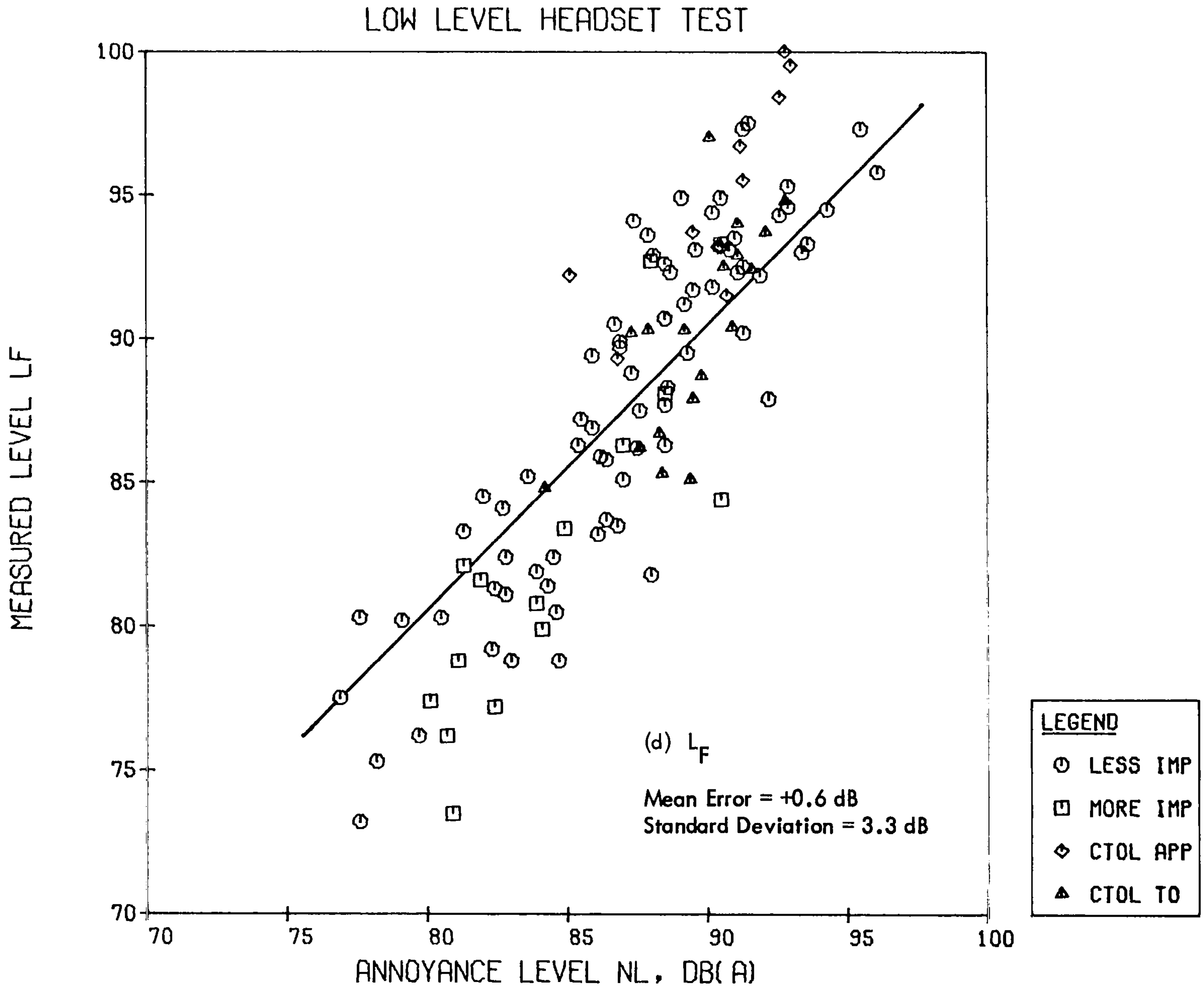


Figure 7. (Continued)

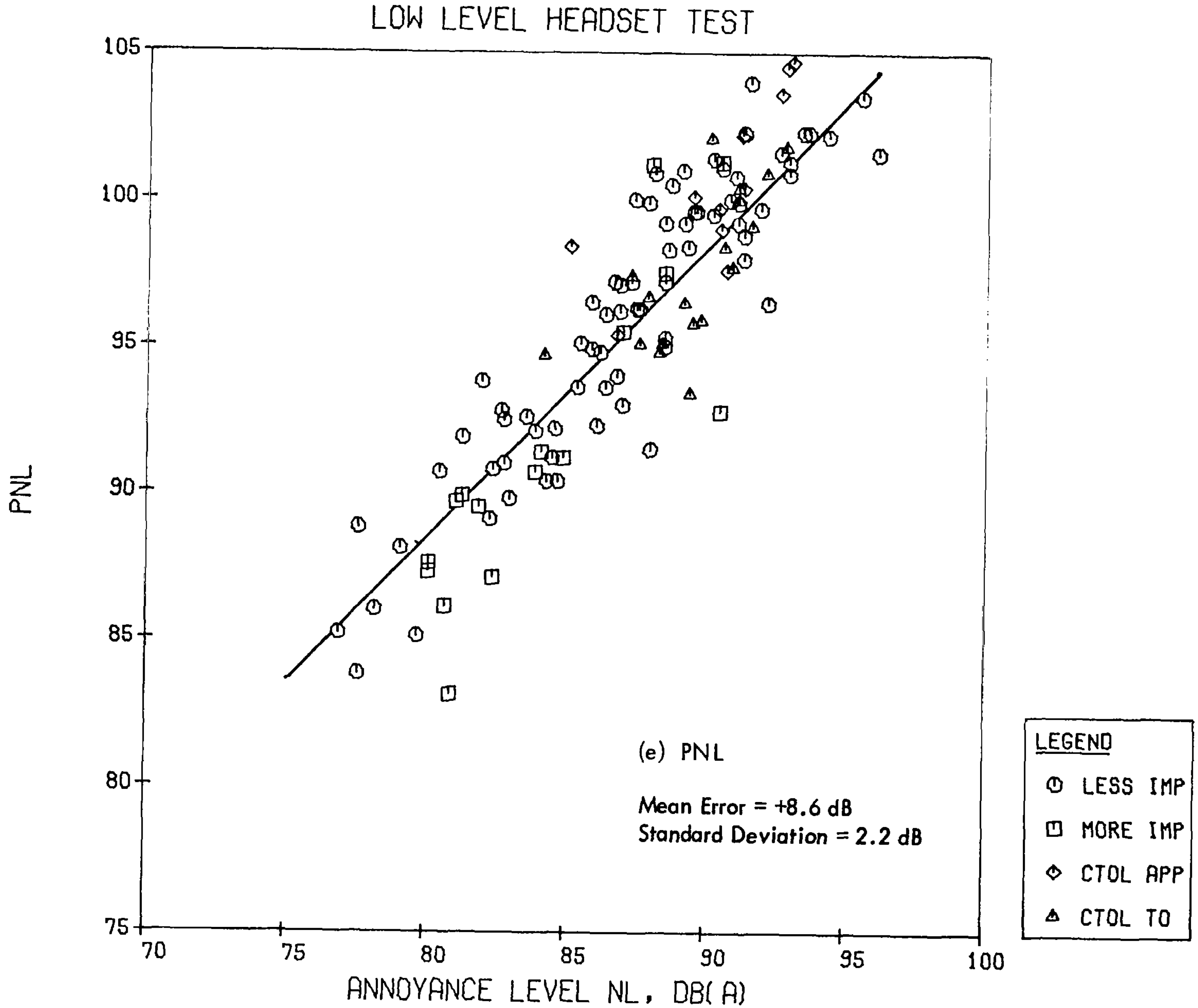


Figure 7. (Continued)

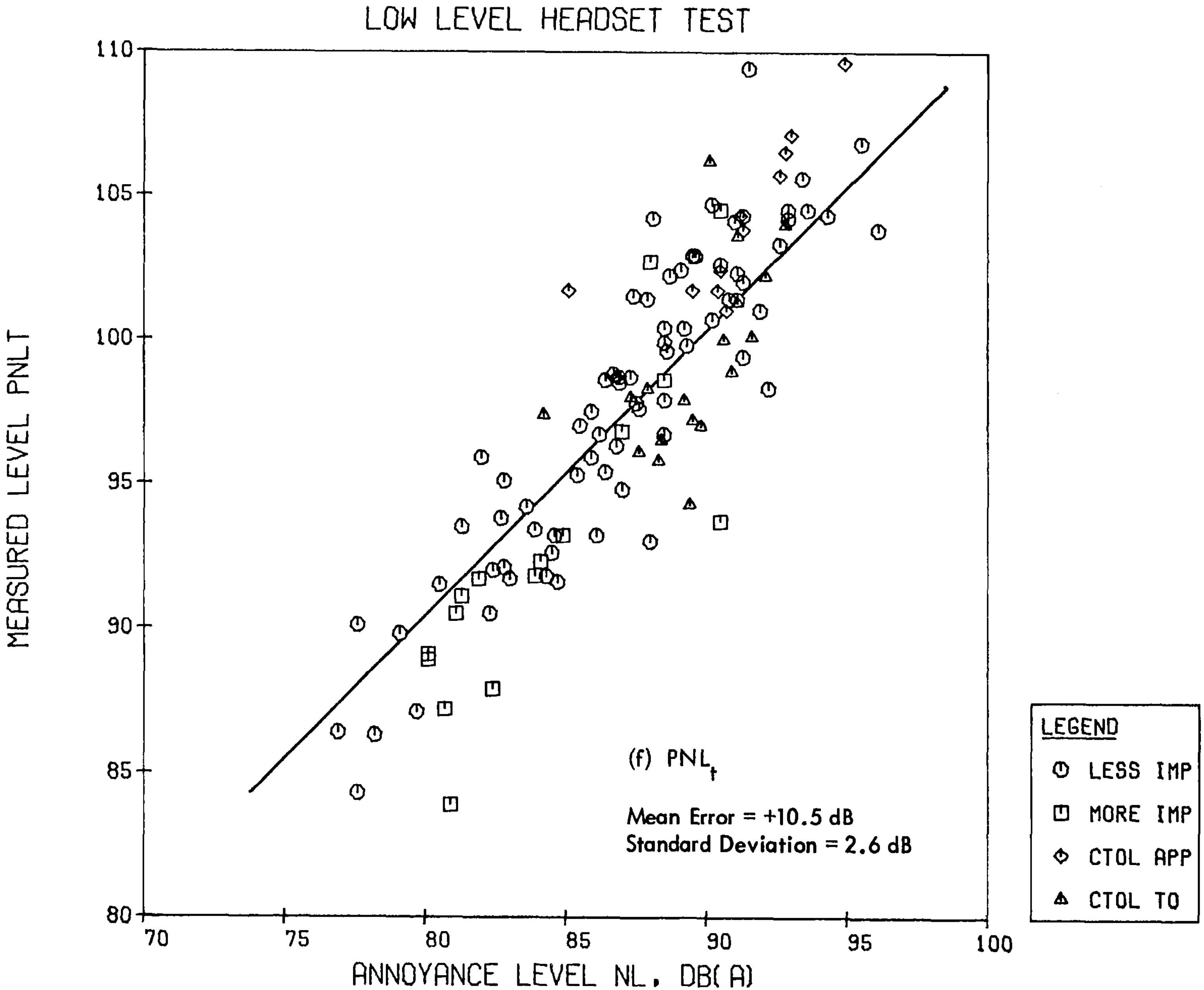


Figure 7. (Continued)

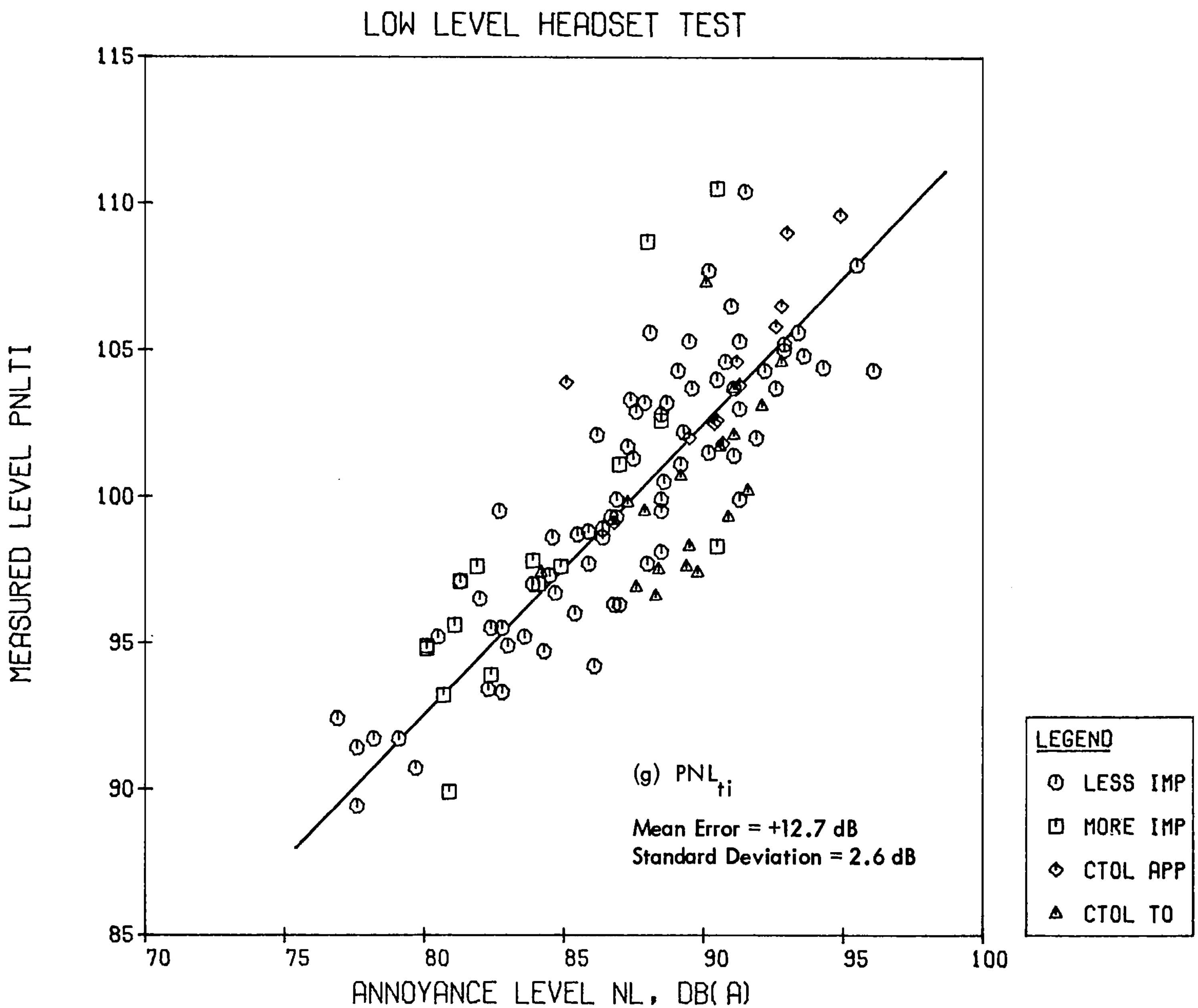


Figure 7. (Continued)

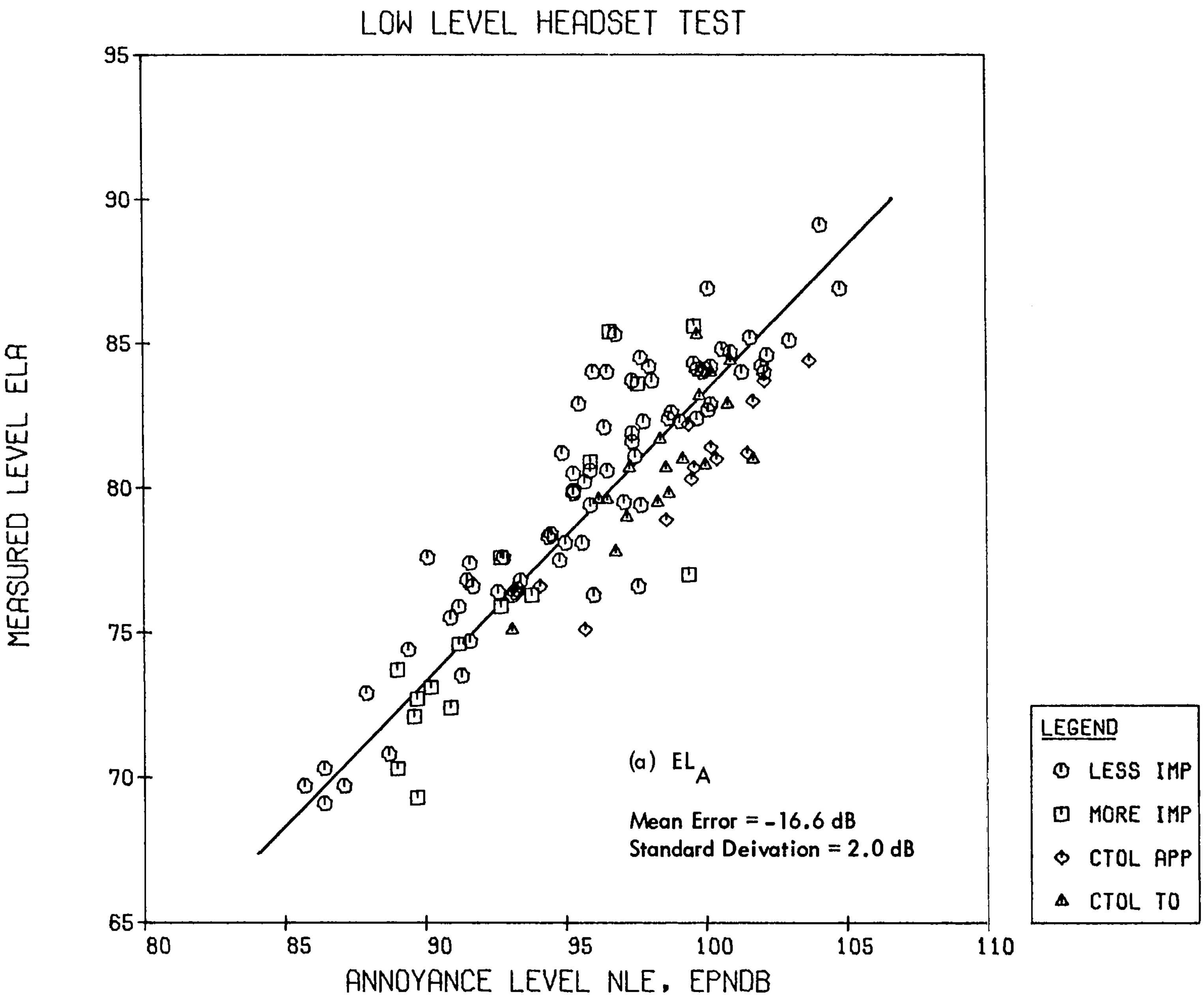


Figure 8. Measured Levels Versus Judged Annoyance Levels; Time-Integrated Scales

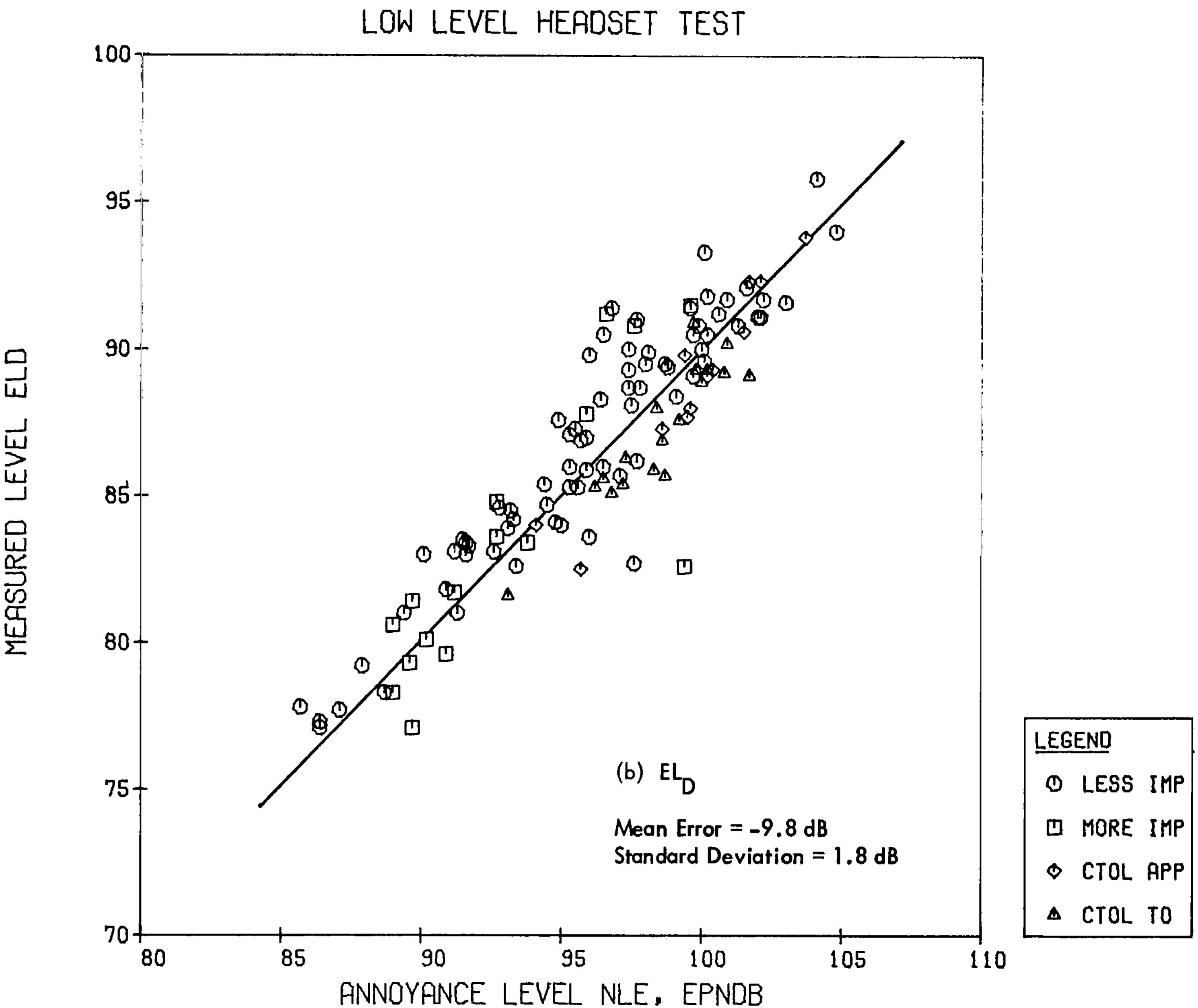


Figure 8. (Continued)

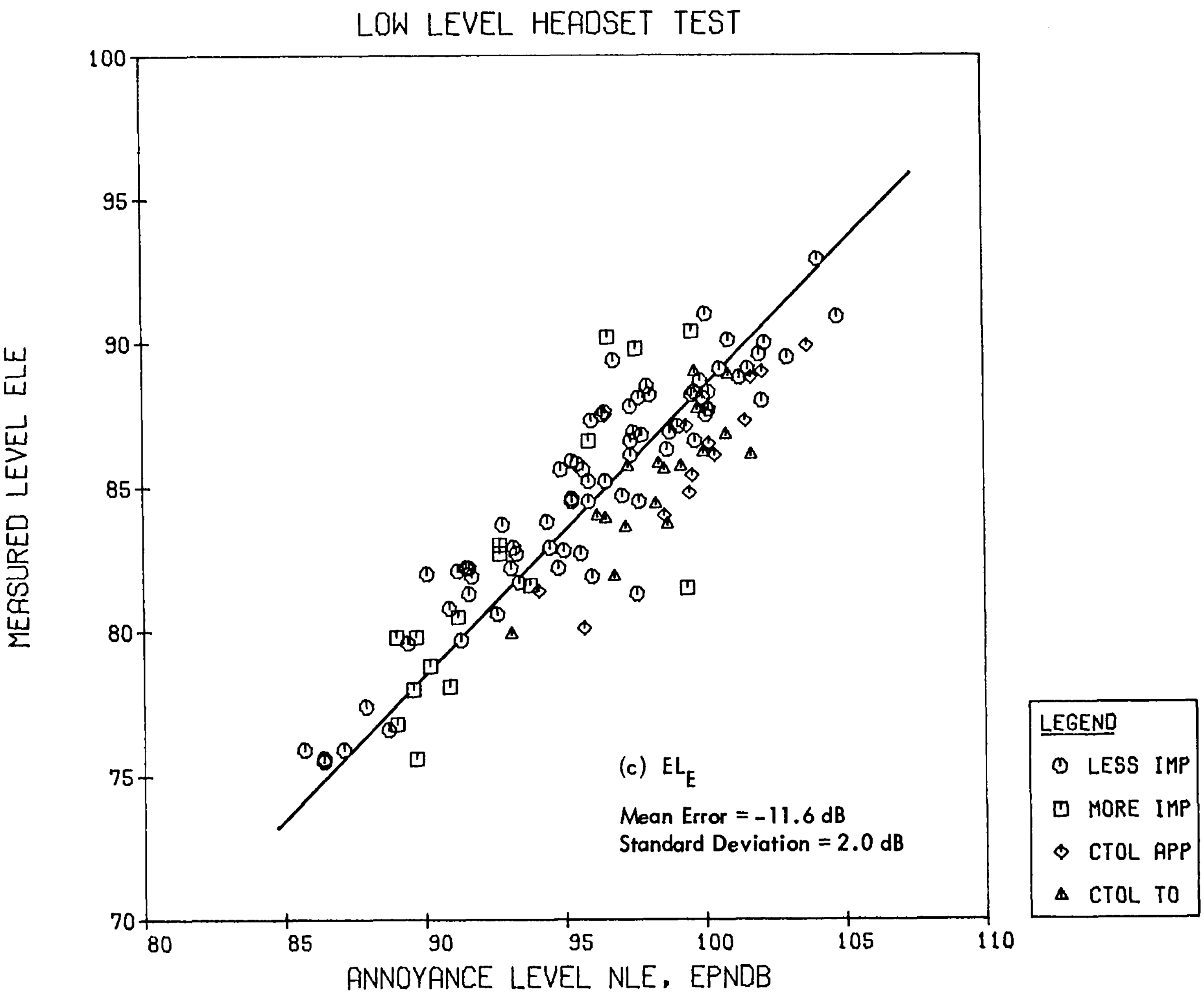


Figure 8. (Continued)

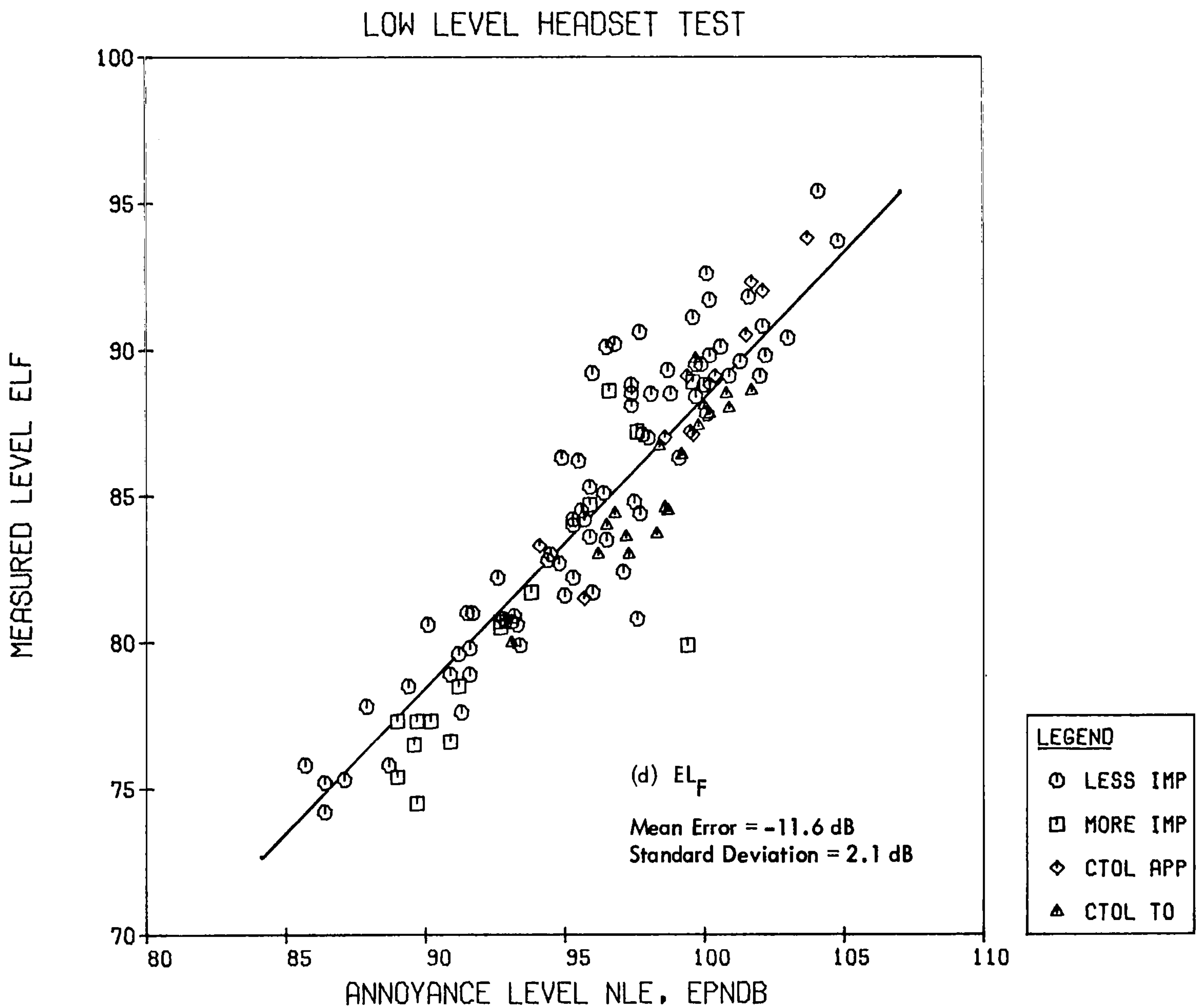


Figure 8. (Continued)

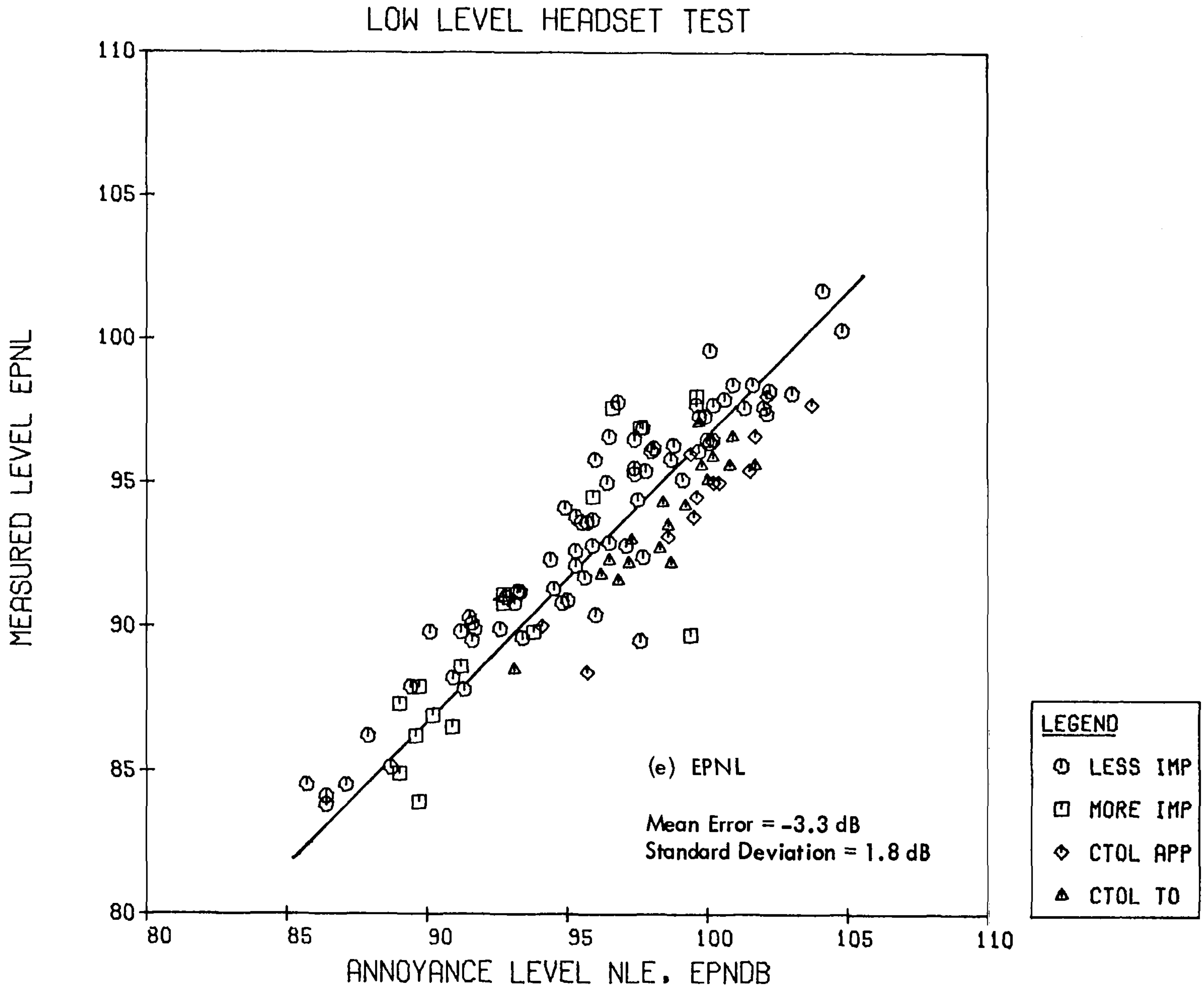


Figure 8. (Continued)

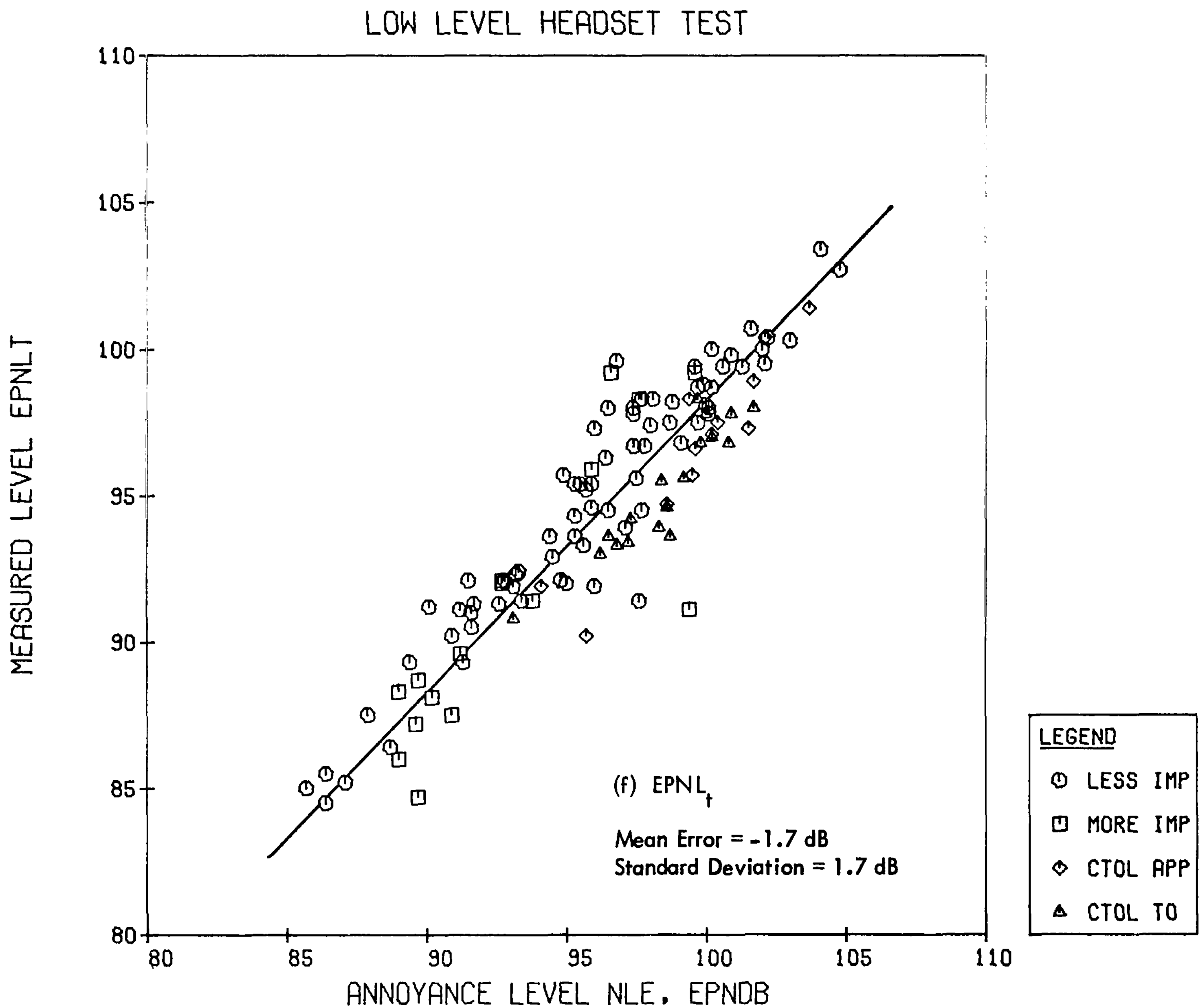


Figure 8. (Continued)

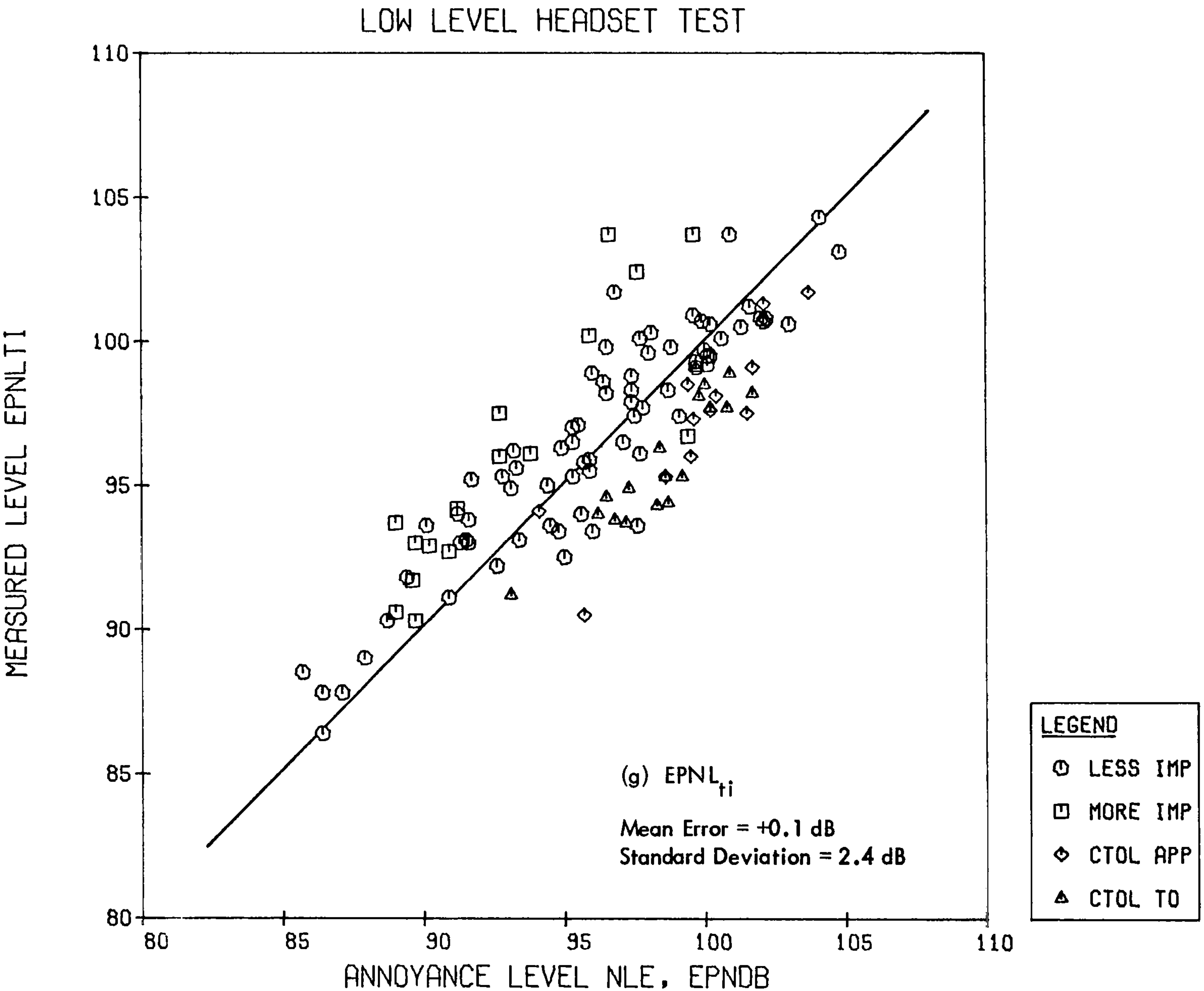


Figure 8. (Continued)

Table 4

Main Experiment Annoyance Prediction Errors, dB
Mean errors for subsamples are relative to overall mean error
(listed for all 119 sounds). Standard deviations in parentheses

	Maximum Levels			Time-Integrated Levels		
Scale	All 119 Sounds	89 Helos / 30 CTOLs	73 less imp. / 16 more imp. / 12 approach / 18 takeoff	All 119 Sounds	89 Helos / 30 CTOLs	73 less imp. / 16 more imp. / 12 approach / 18 takeoff
			+0.2 (2.4)			+0.7 (1.7)
		-0.2 (2.6)	**		+0.5 (1.9)	(**)
			-1.8 (3.2)			-0.1 (2.7)
L_A	-4.6 (2.5)	**		-16.6 (2.0)	** (*)	
			+1.0 (1.8)			-2.4 (1.0)
		+0.3 (1.9)			-1.6 (1.4)	**
			-0.1 (1.9)			-1.0 (1.5)
			+0.1 (2.1)			+0.6 (1.5)
		-0.2 (2.3)	** (**)		+0.5 (1.8)	(**)
			-1.6 (3.0)			+0.2 (2.6)
L_D	+2.3 (2.3)			-9.8 (1.8)	** (**)	
			+2.6 (2.1)			-1.0 (1.1)
		+0.7 (2.6)	*		-1.3 (1.0)	
			-0.6 (2.1)			-1.5 (0.9)
			+0.1 (1.9)			+0.5 (1.6)
		-0.1 (2.1)	(**)		+0.6 (1.8)	(**)
			-0.9 (2.9)			+0.7 (2.7)
L_E	+0.3 (2.1)			-11.6 (2.0)	** (**)	
			+1.3 (2.0)			-2.2 (1.0)
		+0.1 (2.1)	**		-1.8 (1.2)	
			-0.7 (1.8)			-1.6 (1.3)
			0.0 (2.9)			+0.5 (1.9)
		-0.5 (3.1)	**		+0.2 (2.1)	**
			-2.7 (3.2)			-1.0 (2.5)
L_F	+0.6 (3.3)			-11.6 (2.1)	(**)	
			+4.1 (2.1)			-0.4 (1.3)
		+1.7 (3.1)	**		-0.6 (1.4)	**
			0.0 (2.6)			-1.2 (1.1)
			+0.2 (2.0)			+0.9 (1.4)
		-0.1 (2.2)	** (*)		+0.6 (1.7)	(**)
			-1.3 (2.8)			+0.4 (2.4)
PNL	+8.6 (2.2)			-3.3 (1.8)	** (**)	
			+1.6 (1.8)			-2.0 (1.0)
		+0.3 (2.1)	**		-1.7 (1.0)	
			-0.6 (1.9)			-1.5 (0.9)

Table 4 (Continued)

Scale	Maximum Levels			Time-Integrated Levels		
	All 119 Sounds	89 Helos 30 CTOLs	73 less imp. 16 more imp. 12 approach 18 takeoff	All 119 Sounds	89 Helos 30 CTOLs	73 less imp. 16 more imp. 12 approach 18 takeoff
			+0.1 (2.3)			+0.6 (1.3)
		-0.2 (2.6)	** (*)		+0.5 (1.6)	(**)
			-1.9 (3.2)			-0.1 (2.5)
PNL_t	+10.5 (2.6)			-1.7 (1.7)	** (**)	
			+2.5 (1.7)			-1.3 (1.2)
		+0.4 (2.8)			-1.5 (1.0)	
			-0.9 (2.6)			-1.6 (0.9)
			+0.1 (2.2)			+0.5 (1.7)
		+0.3 (2.5)	(*)		+0.1 (2.4)	**
			+1.3 (3.3)			+2.9 (2.2)
PNL_{ti}	+12.7 (2.6)	*		+0.1 (2.4)	** (**)	
			+0.8 (2.1)			-2.5 (1.5)
		-0.9 (2.7)	**		-2.7 (1.2)	
			-2.0 (2.4)			-2.7 (1.0)
			+0.3 (2.3)			+0.5 (1.6)
		+0.2 (2.3)			+0.8 (1.8)	**
			+0.0 (2.4)			+2.1 (2.0)
PNL_{tc}	+13.1 (2.3)			+0.1 (2.2)	** (**)	
			+0.9 (1.9)			-2.4 (1.4)
		-0.6 (2.3)	**		-2.4 (1.2)	
			-1.6 (2.0)			-2.5 (1.0)
			+0.3 (2.5)			+0.6 (2.0)
		+0.3 (2.5)			+0.9 (2.1)	*
			+0.2 (2.5)			+1.9 (2.3)
L_{Ac}	-1.9 (2.4)			-14.8 (2.5)	**	
			-0.4 (2.4)			-3.4 (1.3)
		-0.7 (2.0)			-2.5 (1.7)	*
			-0.8 (2.4)			-1.9 (1.7)

for approaching and departing (takeoff) CTOL aircraft. Table 4 lists these statistics for all measurement scales including those which are not illustrated by scatter diagrams. It should be noted that the mean prediction error for the maximum levels is referenced to the annoyance level NL in dB(A), whereas for the time integrated levels, the reference is NLE in EPNdB (where NLE = NL + 9). The absolute values of these mean errors are of little importance; it is the differences between them which are of interest.

In Table 4, asterisks within pairs of mean values indicate that the difference is statistically significant according to student's T-test (one asterisk for 5 percent significance level, two for 1 percent). Those errors paired without asterisks are not significantly different. Asterisks (in parentheses) between pairs of standard deviation figures (in parentheses) indicate that their respective error variances are significantly different according to a standard F-test (again at the 5 percent or 1 percent level of significance).

On the basis of a broad comparison between the overall error standard deviations for the maximum levels and the time-integrated levels for all sounds, it is clear that for the commonly used scales, the duration correction is generally beneficial in that the consistency with which the scales predict annoyance level is improved. The improvement is significant at the 1 percent level in the cases of L_A, L_D, L_F, and PNL_t and at the 5 percent level for PNL (without tone correction). For L_F, the improvement is very large, doubtlessly because the uncorrected maximum level is a very poor performer.* For L_E, PNL_{ti}, PNL_{tc}, and L_{Ac}, there is no significant change of this group. The maximum level, L_E, is itself a good index of annoyance but the others involve impulsiveness corrections which generally appear to do little to improve the predictive accuracy of the basic scales to which they are applied. Instead, in every case, the impulsiveness corrections counter the beneficial effects of the duration allowance (compare L_A and L_{Ac}, PNL_t and PNL_{ti}, PNL_t and PNL_{tc}).

*The large effect of the duration correction in the case of the F-weighting is possibly linked to a correlation between low frequency energy and duration. Note for example that the improvement is extremely large for the CTOL subsample (standard deviation falls from 3.1 dB to 1.4 dB) for which the takeoffs have longer durations and more low frequency energy than the approaches.

Examination of the subgroup results shows that the non impulse-corrected maximum level scales tend to underestimate the annoyance levels of the more impulsive helicopters relative to the less impulsive ones by around 2 dB. However, this difference nearly vanishes when the duration allowance is included (except in the case of L_F) implying a degree of correlation between impulsiveness and signal duration. Confining attention to the simple weighted sound level scales, it may also be noted that the mean differences between more and less impusive helicopters tend to decrease slightly as emphasis is transferred from high frequencies to low $\left[1.5\,(EL_F) \rightarrow 0.8\,(EL_A) \rightarrow 0.4\,(EL_D) \rightarrow 0.2\,(EL_E)\right]$. This suggests a positive correlation between impulsiveness and low frequency energy in the helicopter sounds.

Many of the subgroup error deviations are considerably smaller than the overall values. This is particularly true of the CTOL sounds (for which the standard deviations are of the same order as the experimental error, i.e., as about as low as could be expected from an ideal noise rating scale). The standard deviations for the helicopters are also small in absolute terms but for all scales except L_{Ac} they are significantly greater than the CTOL values (i.e., practically all scales predict noise annoyance levels less consistently for helicopters than for CTOLs).

Another feature which is common to all duration corrected scales but one (EL_F) is that on average they overestimate annoyance levels of helicopters relative to those of CTOLs by around 2 dB. The F-weighted scale appears to overcome this deficiency by assigning relatively more weight to higher frequency energy than the other scales, thus increasing the relative levels of the CTOL sounds (this is particularly noticeable for the CTOL approach sample).

Turning now to the question of impulse corrections, it is apparent that all the conventional duration-corrected scales predict the annoyance levels of the more impulsive helicopters with rather poor consistency. (In all cases, except EL_F, the error variance is significantly greater than it is for the less impulsive sample at the 1 percent level.) This difference is eliminated for all impulse-corrected scales, whether they involve the ISO factor or the crest factor based term. However, this "improvement" is achieved at least as much by increasing the variance for the less impulsive sample as it is by decreasing the variance for the more impulsive sounds. Consequently, for the impulse-corrected scales, there are increases in the variances for the combined helicopter sample and for the total sample. However, for these scales, the pooled standard deviations for the subgroups are little larger than those of the uncorrected scales; the substantial increases in the overall variances arise because the impulse corrections generate significant differences between the mean prediction errors for the two helicopter subgroups and increase the differences between helicopters and CTOL means.

This is clearly evident in Table 5 which ranks the various duration corrected scales with respect to total error standard deviation but also lists the pooled values. (The differences between the first five scales are not significant at the 5 percent level.)

Table 5

Standard Deviations of Annoyance Prediction Errors, in dB, for Duration-Corrected Annoyance Scales

	Overall Standard Deviation	Pooled Group Standard Deviation
$EPNL_t$	1.7	1.7
EPNL	1.8	1.7
EL_D	1.8	1.8
EL_A	2.0	2.0
EL_E	2.0	2.0
EL_F	2.1	2.1
$EPNL_{tc}$	2.2	1.8
$EPNL_{ti}$	2.4	1.9
EL_{Ac}	2.5	2.2
$EPNL_i$	2.6	2.1

This general review of the performance of the different noise scales begins to reveal the difficulties of isolating the contributions of the various factors such as frequency distribution, tonality, signal duration, and impulsiveness to annoyance, especially when there is a degree of association between them. In general, it seems reasonable to conclude that duration is a most important factor while tonality (as measured by the tone correction in $EPNL_t$) is of minor importance.

The two impulsiveness corrections enhance the consistency with which the noise scales predict annoyance levels of the more impulsive helicopter sounds when they are considered in isolation but, on average, the overall magnitude of the correction is too great, causing the more impulsive helicopters to be overrated with respect to the less impulsive ones. This, together with an increase in error variance for the less impulsive helicopters, causes the disadvantages of the corrections to outweigh their advantages.

To obtain a more quantitative evaluation of the roles of the various underlying factors, it is helpful to turn to multiple regression analysis which yields the coefficient in an optimum annoyance predictor formula comprising a linear combination of the variables.

4.3 Multiple Regression Analysis

The equivalent level $EPNL_{ti}$ of any test sound may be written:

$$EPNL_{ti} = L + D + T + I$$

where	Maximum Level	L	=	PNL
	Duration Correction	D	=	EPNL - PNL
	Tone Correction	T	=	$EPNL_t$ - EPNL
	Impulse Correction	I	=	$EPNL_{ti}$ - $EPNL_t$

The equivalent level is thus a linear combination of these underlying variables but the relative weight attached to each of them is fixed (and equal).

Multiple regression analysis allows the relative weights to vary; the resultant regression analysis gives the best combination. Specifically, it yields the regression coefficients a through e in the linear prediction equation

$$NL' = aL + bD + cT + dI + e$$

The dependent variable NL' is the predicted annoyance level and the regression coefficients are those for which the variance of the prediction errors NL'- NL (predicted annoyance level - actual annoyance level) is minimal. The standard deviation of this error, labeled s_{xy}, is sometimes called the "standard error of estimate."

If the predictor variables are truly independent (uncorrelated), the regression coefficients can be isolated with complete accuracy. However, uncertainty arises when the variables are intercorrelated and in this case the computed regression coefficients have to be assigned a probable error margin (or confidence limits). Table 6 gives the matrix of intervariable correlation coefficients (Pearson's R) for the complete sample of 119 sounds and for the subsamples of 89 helicopters and 30 CTOLs. This shows that the correlation between variables is significant in all cases except (not surprisingly) between impulsiveness and the other variables for the CTOL sample.

The relation between each of these potential predictor variables and annoyance has therefore been examined by a process of "stepwise" multiple regression in which the independent variables are admitted to the analysis one at a time in descending order of importance. At each stage of the analysis, the next most important variable is that which makes the greatest contribution to explained variance. The regression equations defined below exclude variables which were not significant at the 5 percent level.

Table 6

Correlation Matrix for Regression Variables

		D	T	I	$\lvert R_{crit} \rvert$	1% (5%)
	L	-0.647	0.466	-0.616		
All sounds	D		-0.451	0.549	0.235	(0.176)
n = 119	T			-0.434		
	L	-0.610	0.399	-0.586		
All helicopters	D		-0.303	0.505	0.269	(0.205)
n = 89	T			-0.433		
	L	-0.666	0.593	-0.211		
All CTOLs	D		-0.584	0.170	0.449	(0.349)
n = 30	T			-0.257		

For the complete sample of 119 sounds, the regression equation is

$$NL' = 0.92L + 0.56D + 1.1 \qquad (s_{xy} = 1.6\text{ dB}) \ldots \qquad (1)$$

where NL' is the predicted annoyance level and s_{xy} is the standard error of estimate (= standard deviation of residual error NL' - NL). The variables T and I are not significant predictor variables (at the 5 percent level). However, if a dummy variable H is introduced, which takes the value 1 for helicopters and 0 for CTOLs, the result is rather different:

$$NL' = 0.89L + 0.80D + 0.74T - 1.8H + 4.4 \qquad (s_{xy} = 1.4 \text{ dB}) \ldots \qquad (2)$$

The variable T is now significant at the 5 percent level. This result confirms that helicopter sounds are less annoying than CTOL sounds (by an amount equivalent on average to 1.8 dB) and that if this difference is ignored in the predictor model, tone corrections are of little or no value.

If the helicopters (n = 89) and CTOLs (n = 30) are analyzed separately, the two separate regression equations become:

Helicopters: $NL' = 0.89L + 0.78D + 0.90T + 2.63 \qquad (s_{xy} = 1.5 \text{ dB}) \ldots \qquad (3)$

CTOLs: $NL' = 0.89L + 0.73D + 5.4 \qquad (s_{xy} = 0.9 \text{ dB}) \ldots \qquad (4)$

These indicate that the tone correction is an effective annoyance predictor only in the case of the helicopter sounds.

The 95 percent confidence limits for the regression coefficients in the above equations are given in Table 7.

Table 7

Confidence Range for Regression Coefficients

		95% Confidence Range for Regression Coeff. of			
Equation	Sample	L	D	T	H
1	All sounds (119)	0.84-0.99	0.38-0.74	*	**
2	All sounds (119)	0.82-0.97	0.62-0.97	0.22-1.26	-2.5 to -1.1
3	Helicopters (89)	0.81-0.97	0.54-1.02	0.24-1.56	**
4	CTOLs (30)	0.76-1.03	0.53-0.92	*	**

*not significant

**variable not admitted

The large confidence intervals associated with the coefficients of the tone correction term T shows that in those cases where it is a significant predictor variable, it is not a particularly strong one; indeed, in both cases its inclusion reduces the standard error of estimate by a mere 0.05 dB. However, this does not necessarily imply that the tone correction is inappropriate; more probably, it reflects the fact that the term varies very little in this sample of typical aircraft and helicopter sounds (standard deviation = 0.6 dB).

The coefficients of L and D are statistically indistinguishable between the helicopter and CTOL subsamples (Eqs.(3) and (4)); i.e., the regression lines are parallel, separated by the mean difference of around 2 dB. Inclusion of the dummy variable H in the total sample regression (Eq.(2)) thus yields very similar coefficients for L and D. If the variable H is not admitted, the prediction error is significantly greater and the coefficient of D changes markedly (reflecting a degree of correlation between D and H; see Eq.(1)).

Table 7 shows that the coefficients of L and D do not differ substantially from the unit values effectively specified in the $EPNL_t$ formula ($EPNL_t = L + D + T$). Thus, we find in Table 8 that $EPNL_t$ is practically as good an annoyance predictor as the regression equations.

Table 8

Comparison of Annoyance Prediction Errors
$EPNL_t$ vs Regression Model

Sample	Standard Deviation of Error, dB	
	Regression Formula	$EPNL_t$
All sounds (119)	1.6 (1.4*)	1.7
Helicopters (89)	1.5	1.6
CTOLs (30)	0.9	1.0

*including dummy variable H

4.4 Further Analysis of Helicopter Results

A comparison of mean annoyance prediction errors for individual helicopter types reveals significant differences, for example, between the Westland Wessex, the Bell 205, and the Bell OH58A. Some of these differences are illustrated in Figure 9 which compares some mean annoyance prediction errors associated with the time-integrated noise level scales.* Five specific helicopter types are selected: Wessex, S64, Puma, Bell 205, and Bell OH58. The first four of these are drawn from four distinct groups of sounds, each of which can be represented by a typical average one-third octave spectrum shape. These groups are listed in Table 9 and the spectra are shown in Figure 10. The spectra have been drawn by eye as a best fit to a superposition of the individual spectra of all members of the groups. The individual spectra are themselves average values obtained by time integrating each one-third octave band level over its own 10 dB-down duration during the flyover. The relative levels of the four spectra in Figure 10 have been adjusted to ensure that the mean prediction errors for the four groups are correctly related on the EL_A scale (this choice of scale is arbitrary and it does not affect the observations which follow).

Table 9

Groupings of Selected Helicopter Types According to Average Spectrum Shape

(Sample Size in Parentheses)

Group I	Group II	Group III	Group IV
Bell 204 (10)	Squirrel (5)	S76 (5)	Wessex (5)
Bell 205 (4)	Bo 105 (8)	Puma (6)	
Bell 212 (4)	S64 (4)	Super Frelon (5)	
	Bell 206 (3)	S61 (3)	

Progressing from Group I to Group IV, the typical spectra show a progressive shift in energy distribution from low frequencies to high. The Group I helicopters, all members of the large two-blade helicopter family related to the military UH1,

*Because the ISO and crest factor impulse corrections are highly correlated (between ($EPNL_{ti}$ - $EPNL_t$) and ($EPNL_{tc}$ - $EPNL_t$), the correlation coefficient for all 89 helicopter sounds is 0.94), they may be regarded as equivalent measures of the same characteristic.

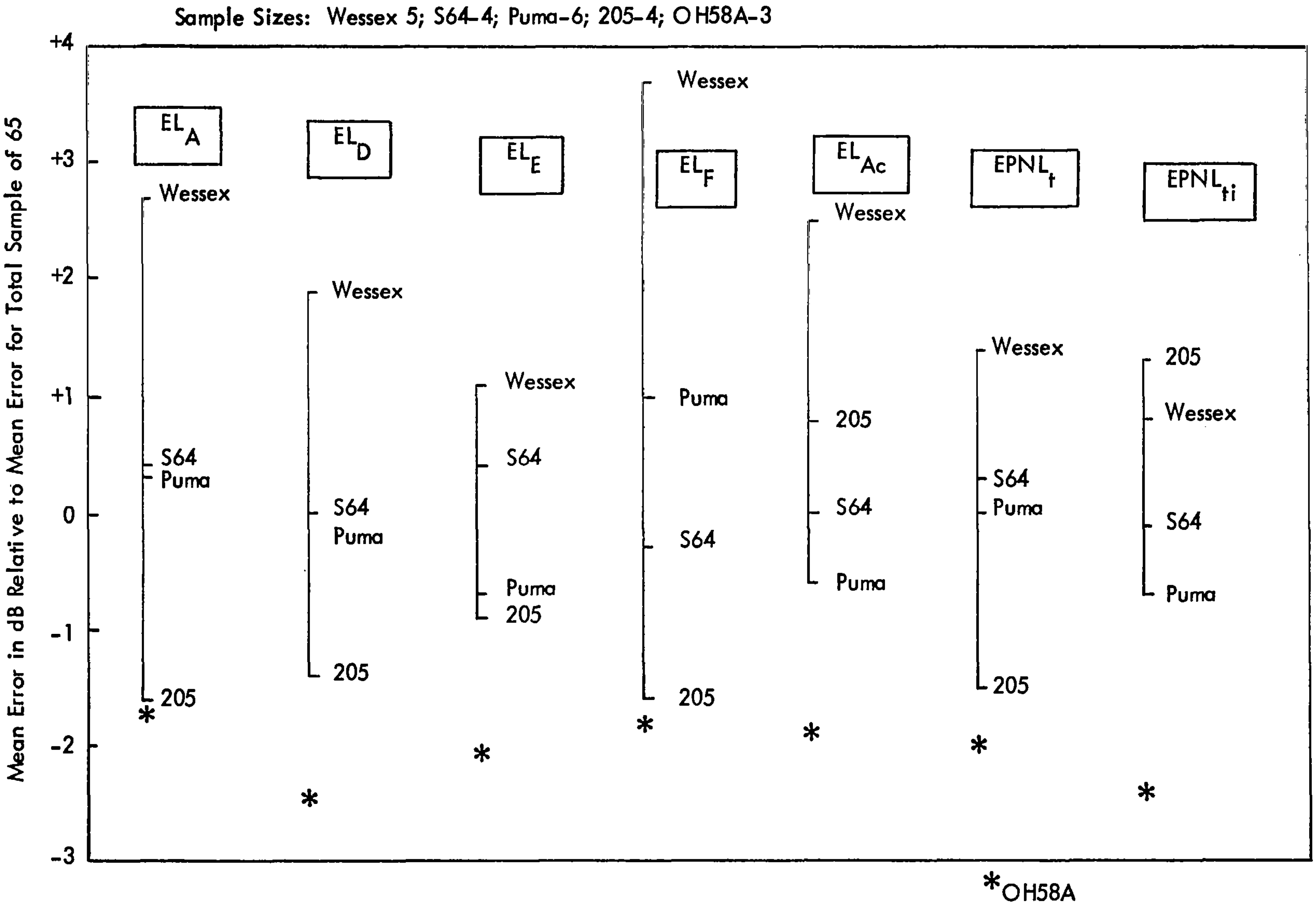


Figure 9. Relative Mean Annoyance Prediction Errors for Selected Helicopter Types

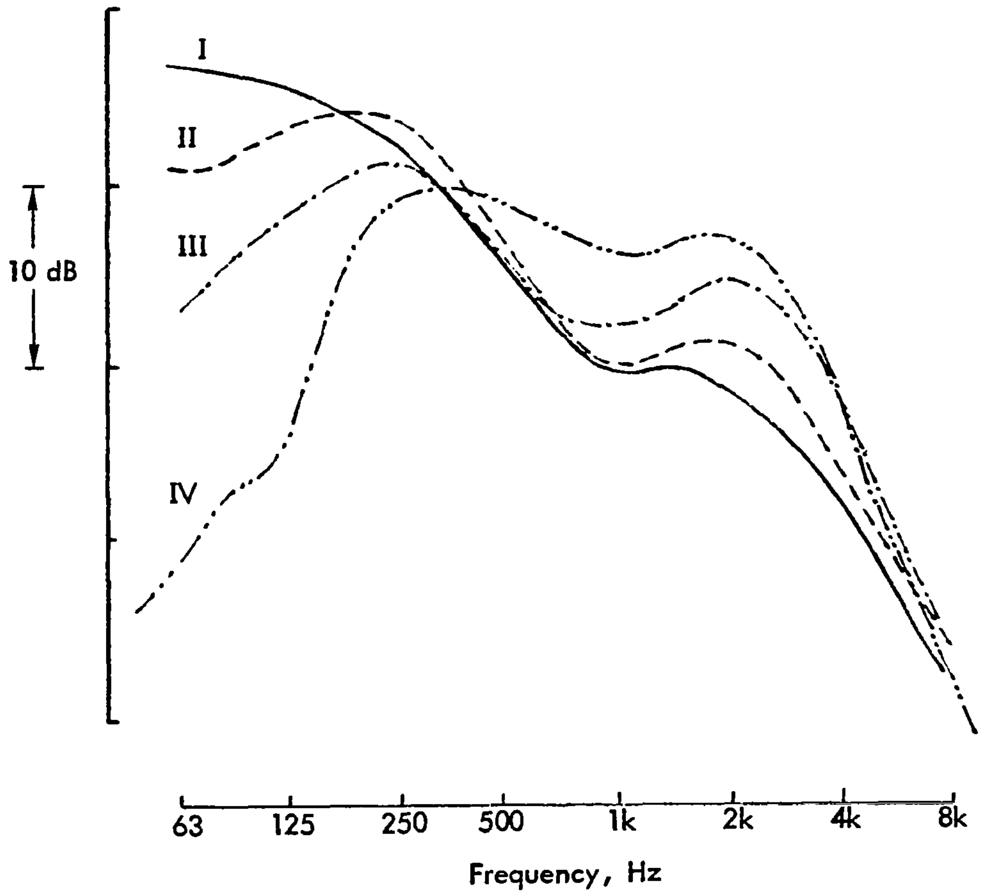


Figure 10. Typical Average Spectra for Helicopter Subgroups (at Equally Annoying Levels)

exhibit pronounced main rotor noise with a low fundamental frequency and, often, a high degree of impulsiveness. Their acoustic energy is clearly concentrated at the low end of the audible frequency range. Group IV comprises the five flyover sounds of the Westland Wessex, a turbine-powered derivative of the four-blade S58. The sound of the Wessex is perhaps best described as "nondescript" with little or no impulsiveness and with no particular sound source dominant. Its frequency spectrum is unique among the helicopters studied in that its energy is spread broadly across frequencies above about 250 Hz with little below that limit.

The OH58A, for which results are also included in Figures 9 and 11, is the military version of the ubiquitous two-blade Bell 206 Jet Ranger. Its spectrum does not fit any of the four groups but it is of special interest because it appears to be a deviant type (in respect of mean annoyance prediction error) and it was one of the two helicopters used in the Wallops Island field experiment[10] (indeed, the recordings used in this study were made during that experiment).

Figure 11 compares the mean annoyance prediction errors, together with their respective 95 percent confidence intervals, for the four groups of sounds. This diagram indicates that of the four sound level weighting functions, "F" is the least appropriate for helicopter noise since it clearly separates the four results. (The differences between the group means are all significant at the 5 percent level.) The A-scale shows some improvement in that the Group II and III errors merge but Group I and IV remain significantly different. For the D-scale, the collapse is more complete with only the Group IV (Wessex) data significantly deviating (at the 5 percent level). No deviations occur in the case of the E-scale for which no differences between means are significant at the 5 percent level.

Figure 12 provides a further comparison of the four frequency weighting functions corresponding to the A, D, E, and "F" scales. Here, the reference levels of these curves have been shifted so that the average levels for all 89 helicopter sounds would be the same on each duration corrected scale. (Thus there is a 6.8 dB difference between the A-curve and the D-curve at 1 kHz. The difference between the A-curve and the "F"-curve is 4.7 dB and between A and E it is 5.1 dB.) Relative to the A-weighting, the other curves give less emphasis to the mid-frequencies (250 to 2,000 Hz) and more to the high frequencies (greater than 2 kHz). Below 250 Hz, the "F"-curve differs little from the A-weighting but the D- and E-functions give considerably more weight. Of the four weightings, the "F"-

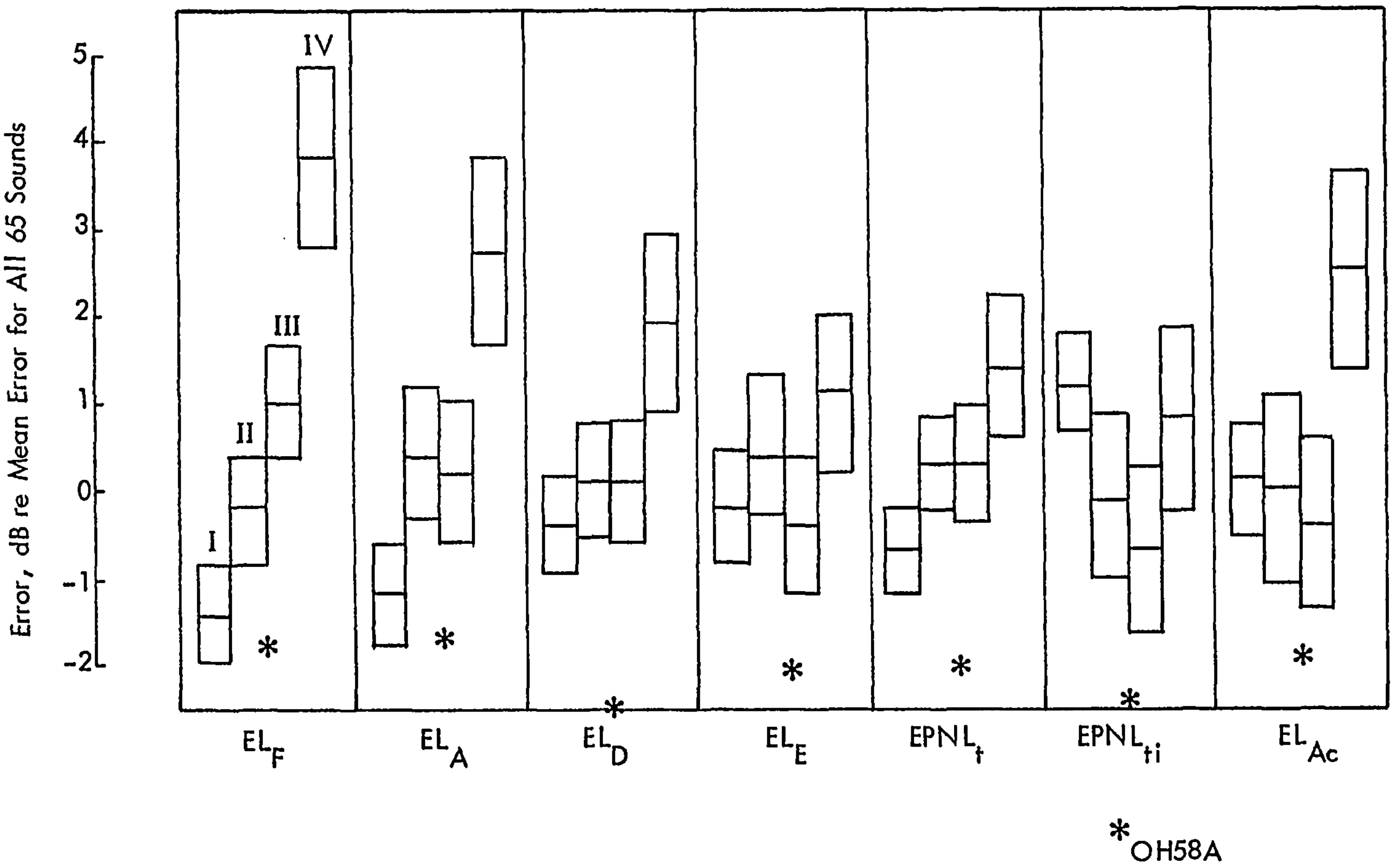


Figure 11. Ninety-Five Percent Confidence Intervals for Helicopter Group Prediction Errors

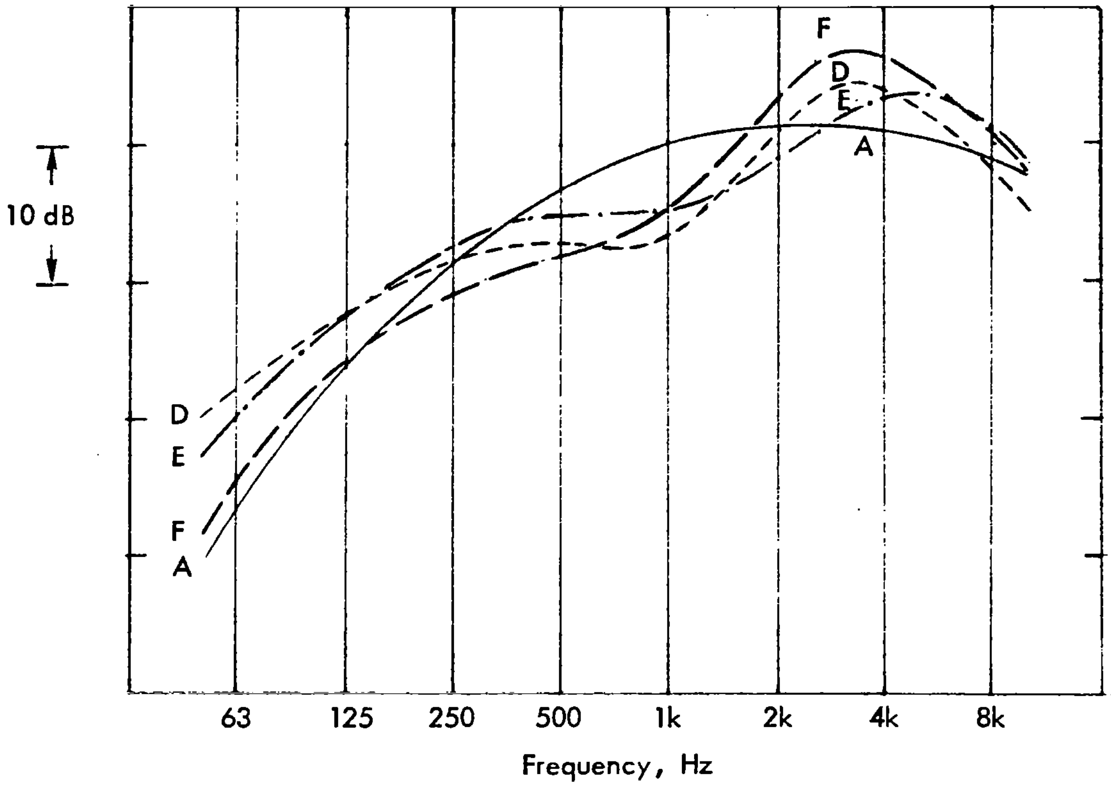


Figure 12. Sound Level Weighting Functions with Relative Levels Adjusted to Give Equal Average Levels for All Helicopter Sounds

curve shows the greatest variation between low frequencies and high. Between 2 and 4 kHz, E and A are similar, "F" applies considerably more weight, and D is intermediate. Above 4 kHz, E becomes dominant but this range is not particularly significant for the helicopter sounds (see Figure 10).

Although it may not be immediately apparent, consideration of Figures 10, 11, and 12 suggests that results for the four groups are harmonized as less weight is given to high frequencies and more to low. It has not been possible to explore this possibility further by fully computing modified sound levels with different weightings from the time histories of one-third octave spectra. However, a realistic assessment of the likely results can be obtained by applying alternative frequency weightings to the time-averaged spectra in Figure 10. Justifications for this approximate procedure may be found in Table 10 where the relative mean prediction errors so calculated are compared with the properly computed values.

Table 10

Comparison of Mean Annoyance Prediction Errors Based On
(a) Full Calculation from Individual One-Third Octave Spectral Time Histories and
(b) Weighting the Typical Average Spectra in Figure 10

	Mean Annoyance Prediction Error, dB									
	EL_A		EL_D		EL_E		EL_F		EL'_D	
Sample	(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)
Group I	-7.0	-7.0	+0.6	+0.2	-0.9	-1.4	-2.6	-3.1	-	-0.3
Group II	-5.7	-5.7	+0.7	+0.9	-0.6	-0.5	-1.7	-1.6	-	+0.5
Group III	-5.6	-5.6	+1.1	+1.0	-1.1	-1.2	-0.2	+0.1	-	-0.7
Group IV	-3.1	-3.1	+2.9	+2.6	+0.4	+0.1	+2.6	+2.0	-	+0.5
Range	3.9	3.9	2.3	2.4	1.5	1.5	5.2	5.1	-	1.2

As noted previously, the simple weighted levels have been normalized by adjusting the overall levels of the four average spectra to equate the two sets of results for the A-scale – the same basis used to determine the equal annoyance levels of the four spectra in Figure 10 (the choice of base scale is arbitrary - the conclusions are unaffected by it).

The agreement in Table 10 between the accurate (a) and approximate (b) methods for applying frequency weighting is good and lends credence to the validity of the figures in the final column of this table which shows further improvement when the D-weighting is slightly modified to reduce the high frequency weighting, i.e., to transfer still more emphasis from high frequencies to low as shown in Figure 13.

Figure 11 indicates that EL_A underestimates the annoyance levels of the Group I sounds but the difference (between Groups I, II, and III) disappears when more weight is assigned to the low frequencies by EL_D. However, the same result is achieved by applying the crest factor impulse correction in EL_{Ac}. The dilemma therefore arises as to whether the Group I sounds (the UH1 family of helicopters) are being underrated because insufficient emphasis is given to low frequency energy or to impulsiveness.* In the case of $EPNL_{ti}$, more weight is given to both factors (than by EL_A) and the Group I sounds are substantially overrated.

The question of impulsivity is considered further in the next section. The analysis in this section has clearly served to illuminate two important general points. The first is that the diagnosis of underlying relationships is hampered by the presence of intercorrelations, even though the test sample is large. The second is that it might be misleading to draw general conclusions from an experiment involving a small number of helicopter types. Figures 9 and 11 indicate, for example, that the Bell OH58A helicopter, which was used for the Wallops Island field tests of the ISO impulse correction is, perhaps, atypical of helicopters in general. On the basis of conventional noise scales, these figures show that, relative to other helicopters, the OH58A is particularly annoying and is thus perhaps an unrepresentative standard by which to gauge them. Had the Wessex been used as a reference aircraft, the case in support of the ISO correction would have been strong (but equally misleading because the Wessex appears to have a particularly inoffensive sound). These results highlight the fact that, as a group, helicopters exhibit a range of acoustic characteristics which is probably greater than for other classes of aircraft and explained why, in general, annoyance levels for helicopter noise are predicted less consistently.

---

*A useful index of the frequency distribution of energy in a sound is the difference between overall (linear) level and A-weighted level, a difference which increases with the concentration of energy at lower frequencies. For all 89 helicopter sounds, the correlation between this index and impulsiveness ($EPNL_{ti}$ - $EPNL_t$) is significant at $R = 0.52$.

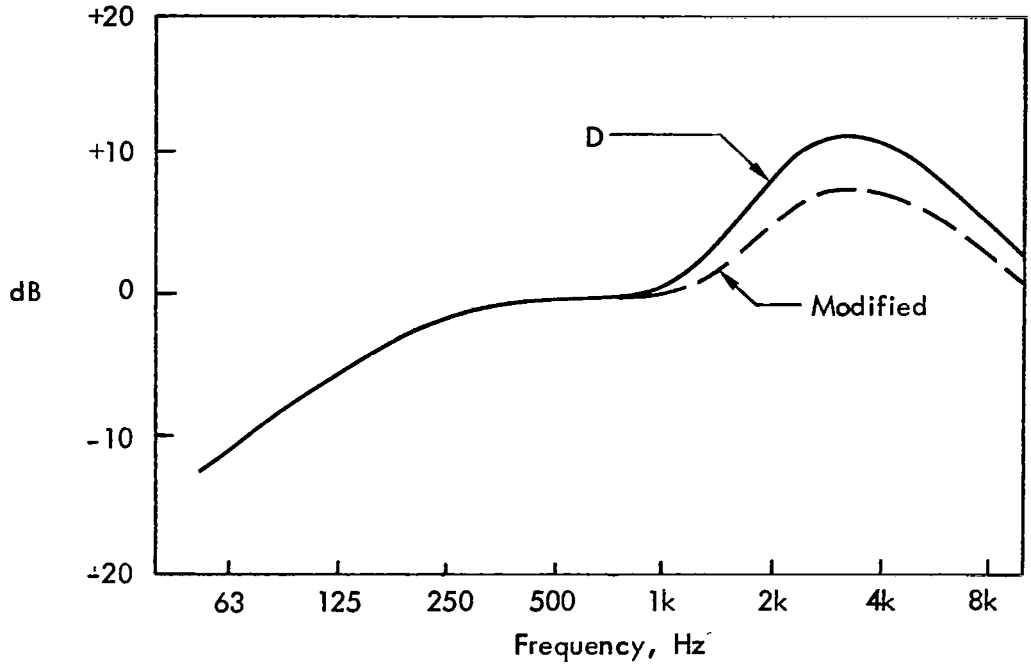


Figure 13. Modified D-Weighting (see Table 10)

4.5 The Need for an Impulse-Correction Term

To some extent, conclusions concerning the appropriateness of the ISO impulse correction are clouded due to the correlation between impulsiveness and low frequency content in the sample of helicopter sounds studied. However, further light may be thrown on the problem by more detailed examination of some individual results.

Table 11 lists the annoyance prediction errors for a subsample of helicopter sounds subdivided by helicopter type. These are the types for which some of the recorded sounds exhibit rather different impulse corrections because recordings were made in both flyover and approach conditions.

This table reveals no tendency for either EL_A or $EPNL_t$ to underestimate the annoyance levels of the more impulsive sounds. Indeed, in the case of EL_A, the converse is true for this particular sample (i.e., it is the less impulsive sounds which are underestimated). There is no significant difference between the two mean errors for $EPNL_t$.

One of the reasons why EL_A and $EPNL_t$ are inherently sensitive to impulsiveness may be deduced from Figure 14 which shows the average one-third octave band spectra* for some of the sounds of Table 11. For each helicopter, the spectra have been overlaid (by eye) so that they coincide at the higher frequencies where the band levels tend to be controlled by noise sources other than the main rotor (i.e., nonimpulsive sources). In all cases, the more impulsive sounds are characterized by significant amplification of spectrum levels in the range 125 to 500 Hz. Since this is the region where weighted band levels of helicopter noise tend to be maximal anyway, impulsiveness directly increases the measured sound levels.

A second factor was evident in Table 4 where, for most of the maximum level scales, there are significant differences between the mean prediction errors for the more and less impulsive helicopter samples (i.e., the maximum measured levels tend to underestimate the judged annoyance of the more impulsive helicopters by a significant amount). This difference largely disappears when duration allowances are included, again suggesting a correlation between impulsiveness and duration.

---

*As in Figure 10, the average one-third octave band levels were computed by time-integration between the 10 dB-down points.

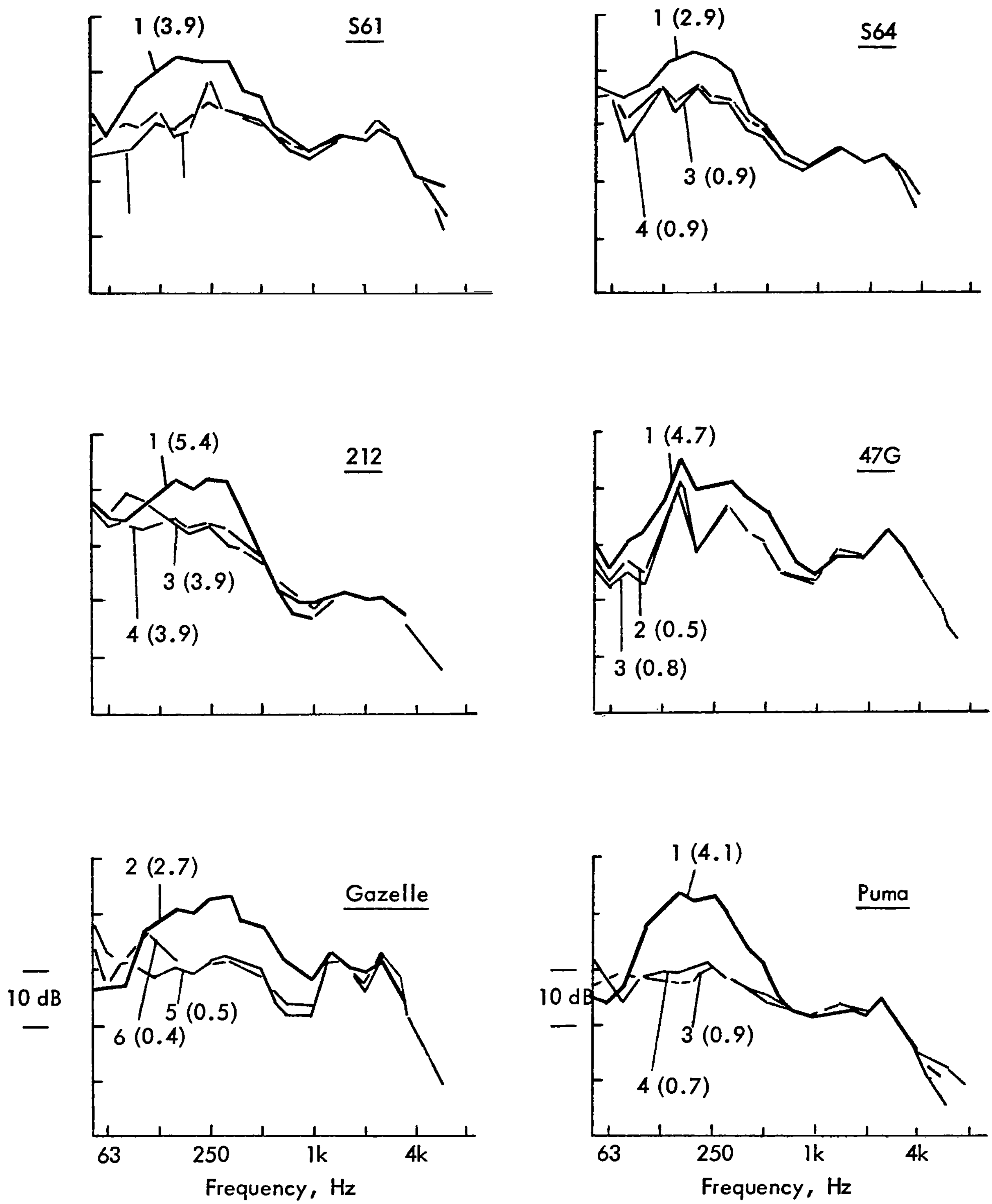


Figure 14. Average Spectra for More and Less Impulsive Recordings of Same Helicopter Types (Impulsiveness corrections $EPNL_{ti} - EPNL_t$ in Parentheses)

Table 11

Annoyance Prediction Errors for Selected Helicopter Sounds, in dB

Sound		D_t †	I †	E_{LA}-NL	$EPNL_t$-NLE
S61	1	-2.9	3.9*	-5.1	-0.4
	2	-5.5	0.7	-7.5	-2.3
	3	-3.9	1.2	-5.5	-0.7
S64	1	-1.0	2.9*	-5.3	-0.1
	3	-4.7	0.9	-6.1	-0.9
	4	-4.4	0.9	-6.5	-1.3
212	1	0.0	5.4*	-5.3	-0.7
	2	-1.3	3.7	-7.8	-2.0
	3	0.0	3.9	-7.9	-2.3
47G	1	-1.4	4.7*	-7.5	-2.4
	2	-1.7	0.5	-5.8	+0.4
	3	-2.3	0.8	-7.3	-1.5
Gazelle	2	-2.2	2.7*	-5.5	-1.0
	3	-3.8	0.8	-6.3	-1.2
	4	-4.6	0.6	-6.0	-0.2
Puma	1	-0.6	4.1*	-4.0	+0.7
	2	-3.4	1.5	-5.3	-0.2
	3	-2.7	0.9	-6.2	-1.3
Mean Errors (Standard deviations in parentheses):					
6 More Impulsive Sounds(*)		-1.4 (1.1)	4.0 (1.0)	-5.5 (1.1)	-0.8 (0.8)
12 Less Impulsive Sounds		-3.2 (1.6)	1.4 (1.2)	-6.5 (0.9)	-1.1 (0.9)

† $D_t = EPNL_t - PNL_t$; $I = EPNL_{ti} - EPNL_t$

The magnitude of such a correlation cannot be measured by computing the direct correlation between duration and impulsiveness without first making allowance for the possibility of sampling bias (e.g., the more impulsive helicopters may have been flying more slowly and thus generating longer signal durations).

The approximate effects of both speed and distance (from the microphone) can, in fact, be eliminated using theory based on spherically symmetric source characteristics. It can readily be shown that in a non-dissipative medium, the duration correction for the sound exposure level of a spherically uniform source passing with speed V at a minimum distance S from an observer increases as $10 \log_{10} (S/V)$. Differences between measured duration corrections D_t (= $EPNL_t$ - PNL_t) not accounted for by this term may therefore be attributed to differences in source directivity. Thus, higher values of the duration increment

$$\Delta = D_t - 10 \log_{10} (S/V)$$

indicate increased sound radiation in forward and/or aft directions.*

Figure 15 shows Δ plotted against the average impulse correction I (= $EPNL_{ti}$ - $EPNL_t$) for the 73 helicopter sounds for which values of S and V are known. A clear correlation between Δ and I is apparent; the correlation coefficient is highly significant ($p \approx 0.001$) at 0.62. This result is totally consistent with the fact that blade slap tends to exhibit pronounced forward directivity. Furthermore, the natural slope of the regression line (0.8) shows that due to impulsiveness, the incremental duration correction approaches the value of the ISO impulse correction.

---

*It is recognized that this analysis involves an oversimplification of actual sound radiation mechanisms. For example, the actual signal durations are also affected by atmospheric sound dissipation which in turn depends upon distance and atmospheric conditions. This "excess attenuation" reduces the signal duration by an amount which increases with the minimum passby distance S of the source. However, in the present case, all but two of the relevant recordings were made at minimum distances no greater than 300 m and, since helicopter noise is dominated by low frequency sound which is less prone to dissipation than high frequency sound, this factor is considered to be of secondary importance; i.e., variations in Δ are largely controlled by variations in fore/aft directivity.

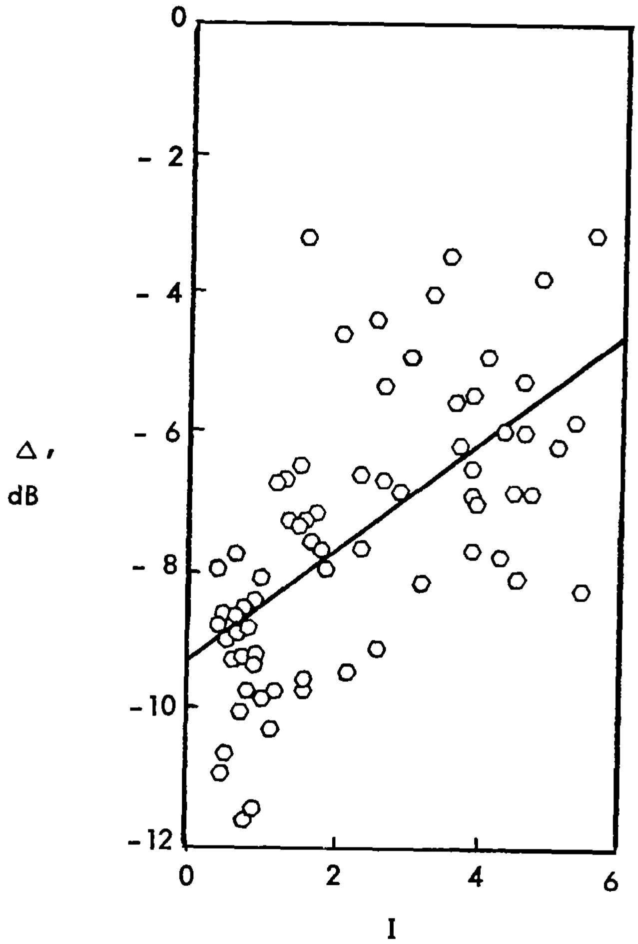


Figure 15. Correlation Between Duration Increment Δ and ISO Impulse Correction Term $I = EPNL_{ti} - EPNL_{t}$

On the basis of the present results alone, it is not possible to state whether it is impulsiveness, low frequencies, duration (or indeed any other correlated variable which may have been overlooked) or some combinations of these which cause increased annoyance. <u>However, if due to the weight of other evidence, the conventional duration allowance made by $EPNL_t$ is accepted a priori together with the standardized frequency weightings, then the results of this study indicate that there is no requirement to include further penalties for impulsiveness.</u>

5.0 DUPLICATE EXPERIMENTS

5.1 Description

The experiments described in Section 4 were repeated using three of the four test tapes, first with the same headphone presentation method described earlier but at a higher average sound level, and second in the acoustic test facilities at NASA Langley Research Center using the two different sound presentation systems of the Interior Effects Room (IER) and the Exterior Effects Room (EER). In the IER, subjects are seated in a typical American living room environment while the test sounds are played through loudspeakers located outside the room. The sound transmission characteristics of the structure are such that the sounds are heard much as they would be inside a typical home. In the EER, subjects are seated inside an acoustically treated lecture theatre and the sounds are played through sets of loudspeakers installed in the ceiling.

A total of 80 subjects took part in the tests at NASA. These were paid volunteers recruited from the general public living in the neighborhood of LaRC. Of the subjects, 60 were female of average age 36.3 years (standard deviation 13.0 years) and 20 were male, average age 28.7 years (standard deviation 11.0 years). They were divided evenly on the basis of age and sex between the two experiments (which were conducted simultaneously and in each of which they participated five at a time).

The signals were measured by microphones located in the middle of the test groups and analyzed on-line. The 1/2-second one-third octave band spectra and impulse corrections were filed on computer discs and returned to Loughborough for subsequent processing and calculation of the various measures of sound level. Table 12 lists summary statistics of the sound levels for all tests, including the main headphone tests (HSLO) described in Section 4.

The only significant difference between the two headphone tests was the level difference of approximately 14 dB (which is measured on all scales). The average levels in the EER were marginally lower than those of HSLO but the range of levels, as reflected in the standard deviations, is also smaller. This is a consequence of the significantly different frequency response of the EER sound replay system which is illustrated in Figure 16 in relation to that of the average headphone. (The standard deviations in Figure 16 are based on LaRC calibration data for the EER; they have been computed from measurements at six locations surrounding the five seating positions used in these tests.)

Table 12

Summary of Test Conditions for Four Experiments

	HSLO (Main) n = 119(1)		HSHI n = 89		IER n = 93		EER n = 93	
	mean(2)	s.d.(3)	mean	s.d.	mean	s.d.	mean	s.d.
NL	87.5	4.3	103.3	5.3	64.6	2.8	85.0	3.9
NEL	96.4	4.3	112.5	5.3	76.4	2.2	93.1	3.4
L	89.9	3.1	103.9	2.6	75.7	2.2	89.0	3.7
L_A	82.8	5.4	97.0	5.2	63.7	3.3	82.2	4.4
L_D	89.8	5.4	104.0	5.5	70.2	2.9	87.7	3.9
L_E	87.8	4.8	102.0	4.8	-	-	-	-
L_F	88.1	6.4	102.4	6.4	67.3	2.5	85.0	4.3
PNL	96.1	5.1	110.4	4.9	77.0	2.8	94.5	3.6
PNL_t	97.9	5.6	112.2	5.4	78.5	3.0	96.0	3.9
PNL_{ti}	100.2	4.6	114.4	4.4	78.6	3.0	96.4	3.7
EL_A	79.8	4.4	93.8	4.1	62.6	2.3	79.5	3.4
EL_D	86.6	4.2	100.7	4.1	69.4	2.0	85.1	3.0
EL_E	84.7	3.9	98.8	3.8	-	-	-	-
EL_F	84.8	5.0	98.8	4.8	67.9	1.7	82.9	2.7
EPNL	93.1	4.0	107.4	3.7	76.2	2.0	92.1	2.8
$EPNL_t$	94.6	4.3	109.0	3.9	77.0	2.2	93.4	2.9
$EPNL_{ti}$	96.5	3.7	110.8	3.4	77.2	0.2	93.7	3.0
D(4)	-3.0	2.2	-3.0	2.3	-0.8	2.7	-2.3	1.3
T(5)	1.5	0.6	1.5	0.5	0.8	0.4	1.3	0.4
I(6)	1.8	1.5	1.8	1.6	0.2	0.4	0.3	0.3

(1) Number of Test Sounds.

(2) Mean level, in decibels, for all sounds.

(3) Standard Deviation, in decibels, for all sounds.

(4) D = EPNL - PNL.

(5) T = $EPNL_t$ - EPNL.

(6) I = $EPNL_{ti}$ - $EPNL_t$.

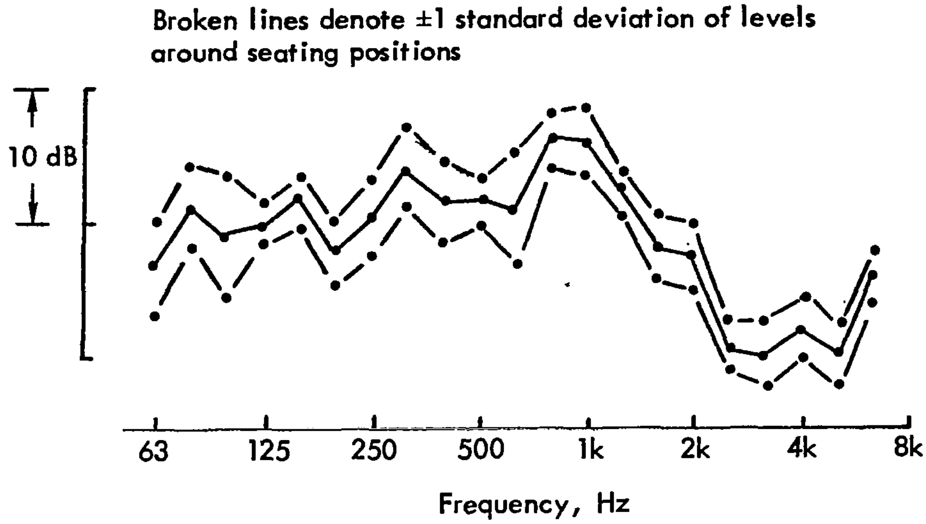


Figure 16. Frequency Response of EER Relative to That of Average Headphones

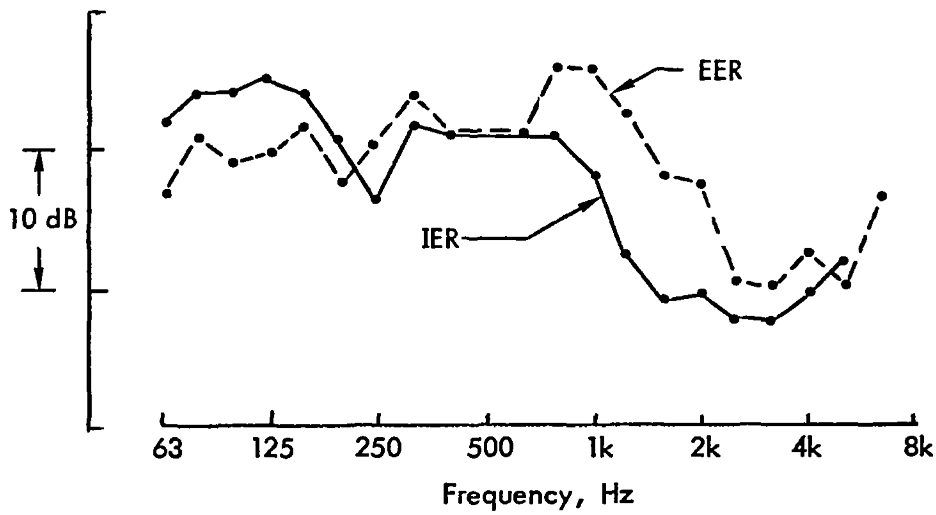


Figure 17. Frequency Response of IER Relative to That of Average Headphones

The levels in the IER experiment were some 17 or 18 dB lower and the variations in level were only about half as great as those in HSLO. This is partly due to even greater differences in the frequency response (Figure 17) but also in the case of the duration-corrected measures because the smaller signal-to-background-noise level in the IER tends to prolong the effective duration of the lower level signals.*

These differences are summarized in Table 13 which compares means and standard deviations of level, duration, tone, and impulsiveness variables for the four tests – for all sounds and for separated helicopters and CTOL samples.

Table 13 clearly shows that a major difference between the headphone and loudspeaker tests is associated with the magnitude of the impulse correction term; in the loudspeaker tests, impulsiveness has all but disappeared. This is presumed to be a consequence of both reverberation and the use of multiple loudspeakers. The average tone correction is also smaller, particularly in the IER.

5.2 Comparison of Results

Example scatter diagrams of the duplicate test results are shown in Figures 18, 19, and 20 (for L_A and $EPNL_t$) but more complete statistics are listed in Tables 14, 15, and 16, which correspond to Table 4 for the main low level headphone experiment. As before, the mean annoyance prediction error listed for all sounds is the average amount by which the measured levels of the test sounds exceed the measured levels of the reference sound when the latter is judged to be equally annoying. The values of these means depend, of course, upon the choice of reference sound and the fact that they differ between tests is partly due to the different characteristics of that sound as heard in the various experiments. The subgroup errors on the other hand are expressed in relation to the overall errors and therefore they are largely independent of the reference sound.

Considering initially the two headphone tests, Figure 21 is a combined scatter diagram for $EPNL_t$. Careful inspection of this figure reveals a nonlinearity in that the swathe of data points curves to the right towards the upper end of the level range. This implies that, on average, the annoyance levels of the test sounds increase more rapidly with level than those of the reference sound. However, this applies more strongly to the CTOL sounds than it does to the helicopter sounds.

*It should be noted that Figures 16 and 17 do not make allowance for the spectral differences between the free field sound (i.e., as measured in the NASA test chambers) and that at the entrance to the ear canal caused by the presence of the subjects' heads.

Table 13

Mean Values (and Standard Deviations), in decibels, of Level, Duration, Tone, and Impulse Variables in Four Tests

	Test	PNL		D		T		I	
All Sounds	Main	96.1	(5.1)	-3.0	(2.2)	1.5	(0.6)	1.8	(1.5)
	HSHI	110.4	(4.9)	-3.0	(2.3)	1.5	(0.5)	1.8	(1.6)
	IER	77.0	(2.8)	-0.8	(2.7)	0.8	(0.4)	0.2	(0.4)
	EER	94.5	(3.6)	-2.3	(2.5)	1.3	(0.4)	0.3	(0.3)
All Helicopters	Main	95.1	(5.2)	-2.4	(1.7)	1.5	(0.5)	2.2	(1.5)
	HSHI	108.6	(4.7)	-2.1	(1.7)	1.4	(0.4)	2.4	(1.6)
	IER	76.4	(3.1)	-0.1	(2.8)	1.0	(0.4)	0.2	(0.4)
	EER	93.4	(3.9)	-1.5	(2.3)	1.3	(0.4)	0.3	(0.3)
All CTOLs	Main	98.9	(3.3)	-4.9	(2.3)	1.8	(0.6)	0.6	(0.4)
	HSHI	114.0	(3.2)	-4.8	(2.3)	1.7	(0.6)	0.7	(0.4)
	IER	78.3	(1.3)	-2.5	(1.7)	0.6	(0.5)	0.2	(0.4)
	EER	96.6	(1.7)	-4.0	(2.1)	1.3	(0.5)	0.2	(0.1)
Less Impulsive Helicopters	Main	95.9	(4.9)	-2.6	(1.7)	1.6	(0.5)	1.5	(0.6)
	HSHI	109.5	(4.3)	-2.3	(1.6)	1.5	(0.4)	1.7	(0.9)
	IER	76.9	(2.8)	-0.7	(2.9)	1.0	(0.4)	0.2	(0.4)
	EER	93.9	(3.6)	-1.8	(2.4)	1.4	(0.4)	0.2	(0.1)
More Impulsive Helicopters*	Main	91.4	(5.2)	-1.3	(1.6)	1.2	(0.3)	4.8	(0.5)
	HSHI	105.6	(4.9)	-0.9	(1.6)	1.2	(0.3)	4.8	(0.5)
	IER	74.2	(3.1)	1.4	(1.8)	0.9	(0.3)	0.2	(0.3)
	EER	91.0	(4.0)	-0.5	(1.6)	1.1	(0.3)	0.6	(0.5)
CTOL Approach	Main	100.9	(3.2)	-6.5	(1.8)	2.2	(0.5)	0.6	(0.6)
	HSHI	116.0	(3.2)	-6.4	(1.8)	2.2	(0.6)	0.6	(0.5)
	IER	77.7	(0.8)	-3.5	(1.3)	0.4	(0.6)	0.0	(0.1)
	EER	96.6	(1.6)	-5.7	(1.7)	1.4	(0.7)	0.2	(0.1)
CTOL Takeoff	Main	97.6	(2.7)	-3.8	(2.0)	1.5	(0.5)	0.6	(0.2)
	HSHI	112.7	(2.5)	-3.7	(2.0)	1.4	(0.3)	0.8	(0.3)
	IER	78.7	(1.4)	-1.8	(1.6)	0.7	(0.4)	0.3	(0.5)
	EER	96.6	(1.8)	-2.9	(1.5)	1.2	(0.3)	0.2	(0.1)

*More Impulsive Sounds are those with $EPNL_{ti} - EPNL_{t} \geq 4$ in Main test.

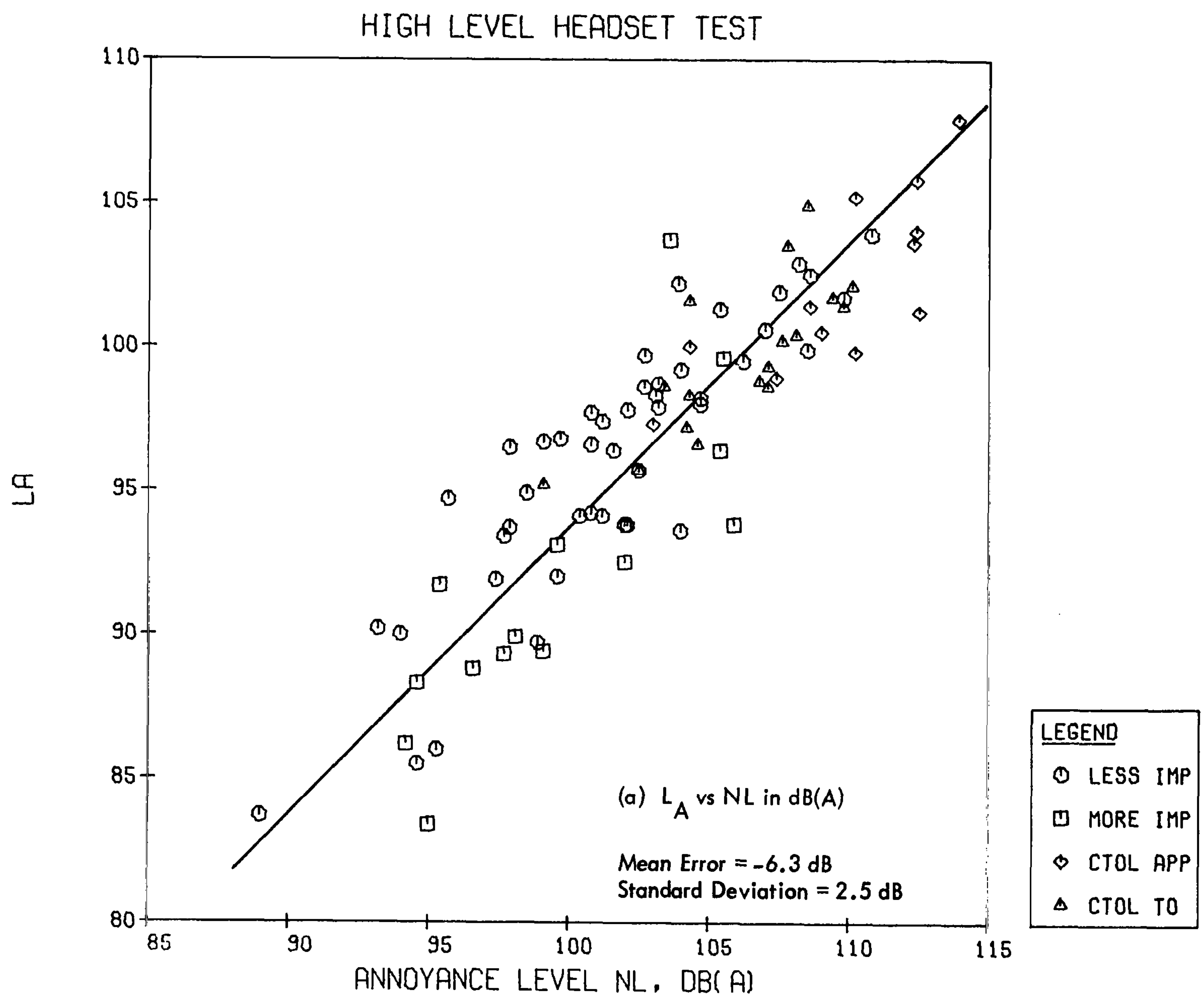


Figure 18. Measured Level Versus Annoyance Level; High Level Headphone Test

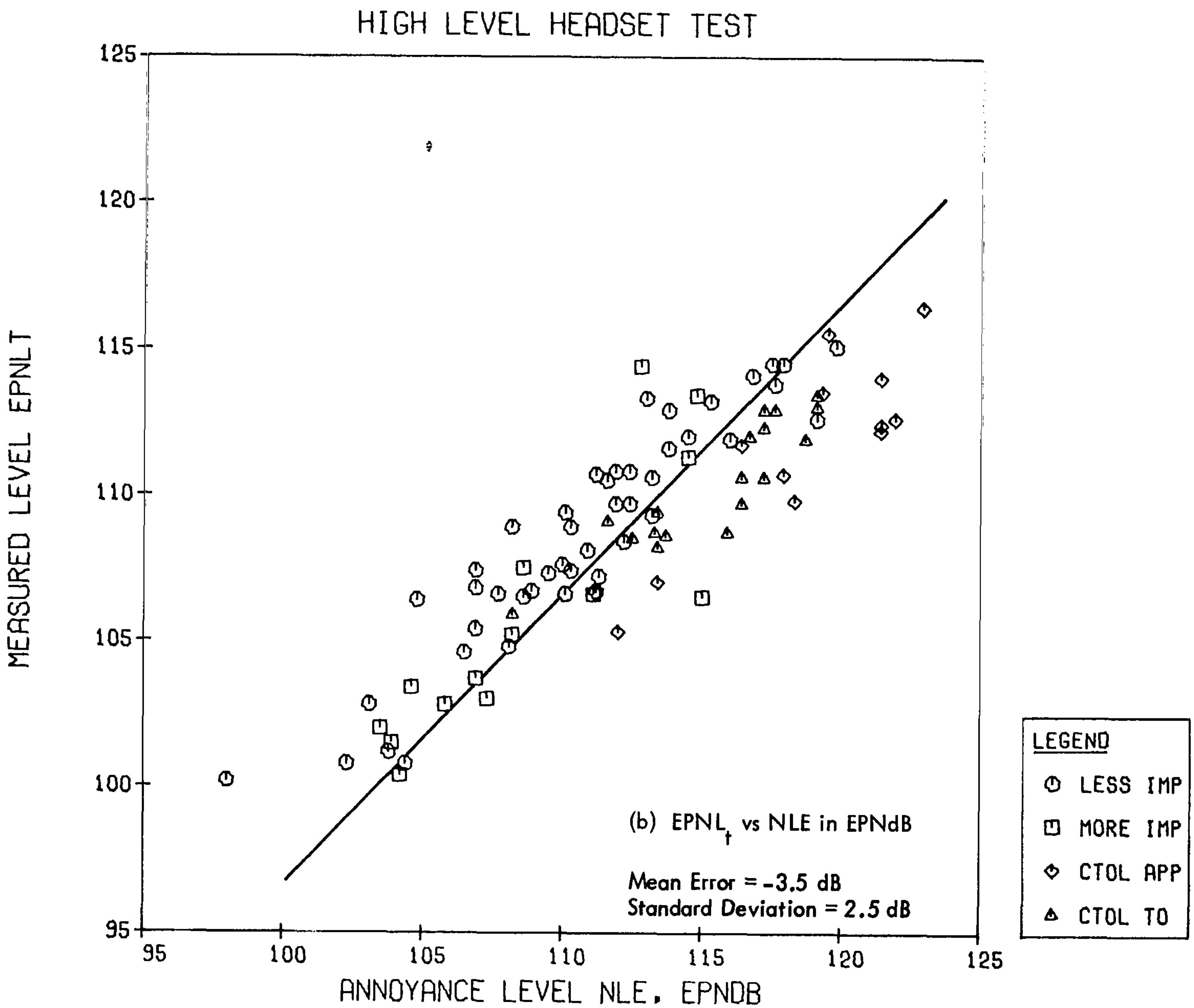


Figure 18. (Continued)

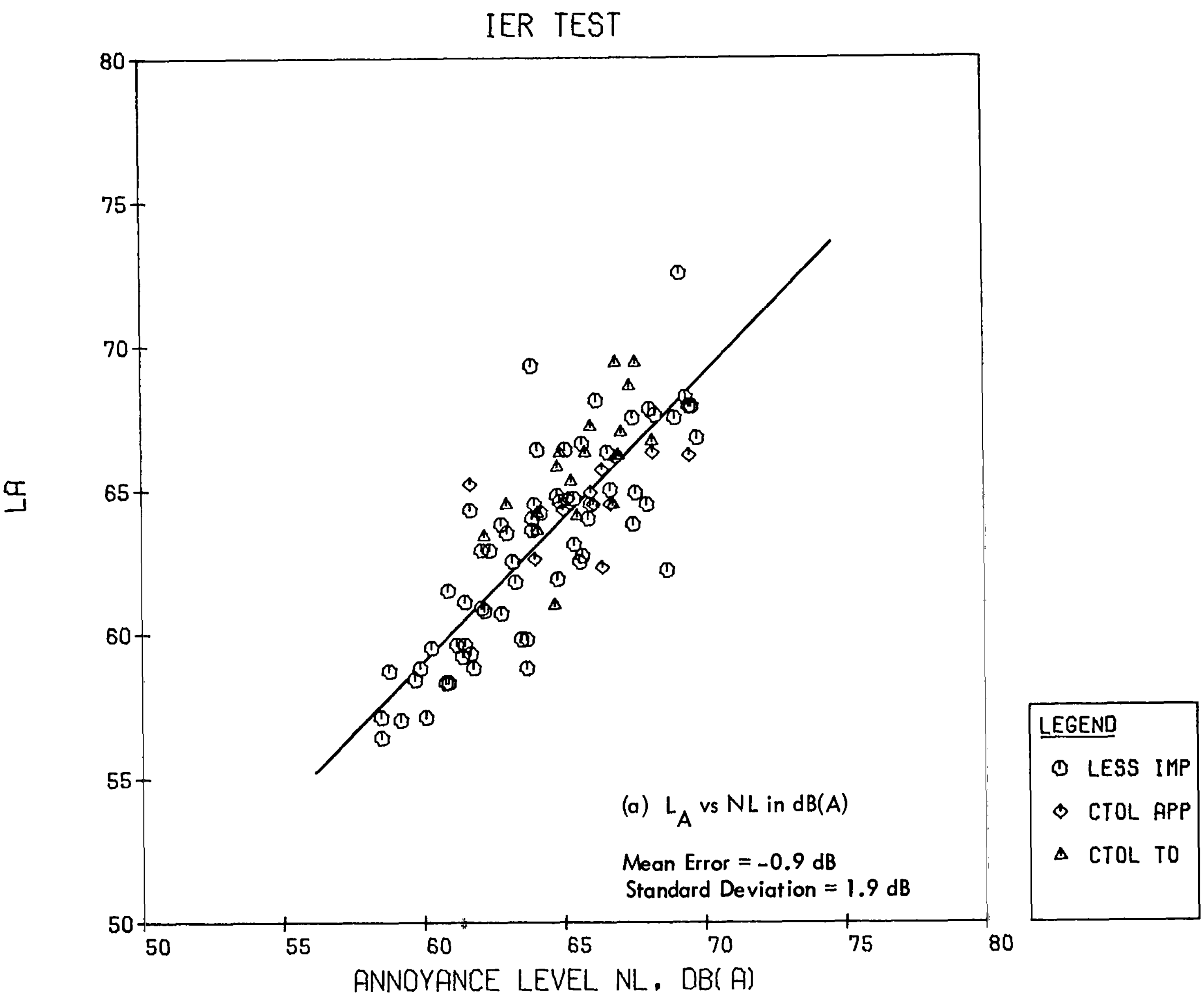


Figure 19. Measured Level Versus Annoyance Level; Interior Effects Room

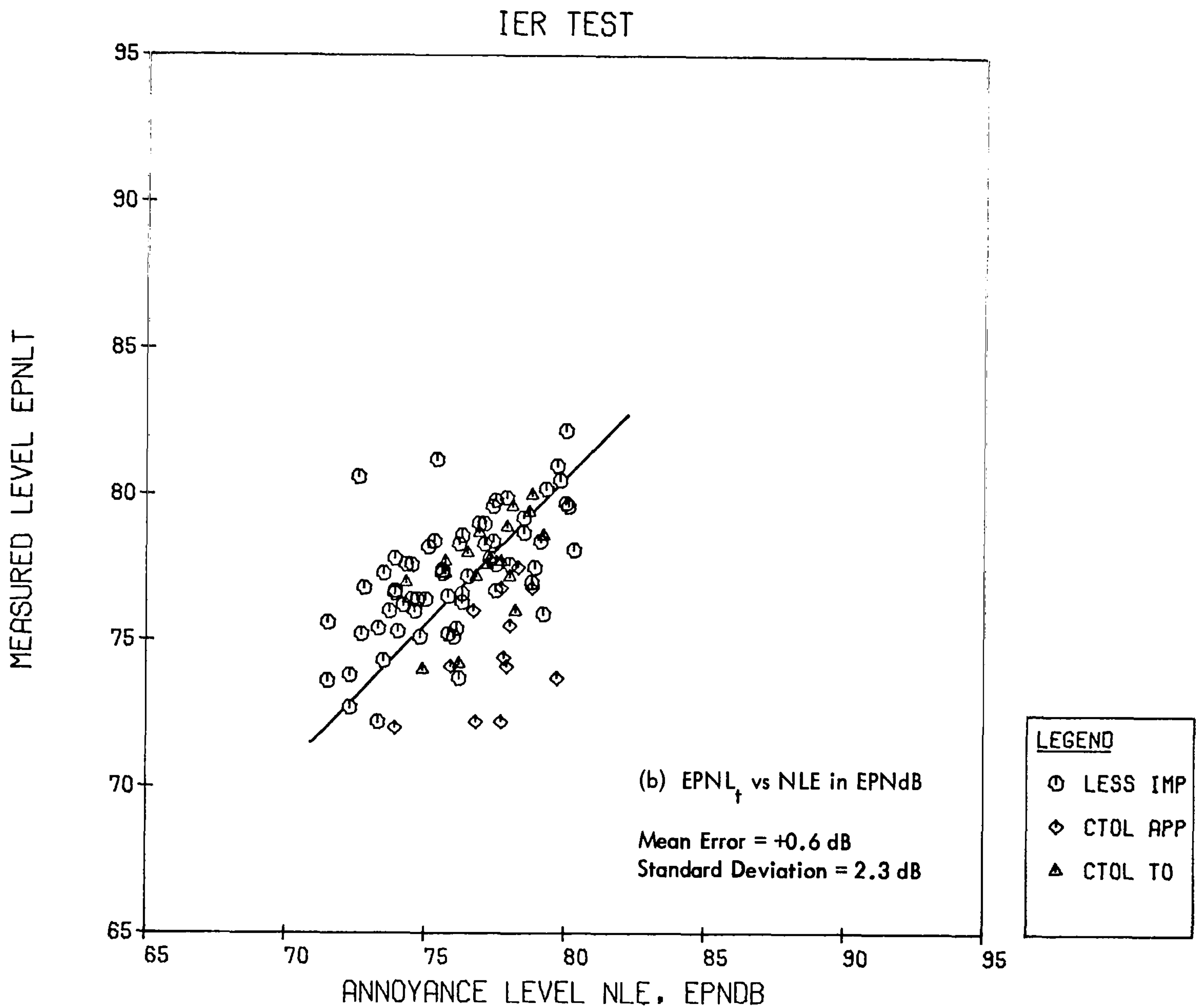


Figure 19. (Continued)

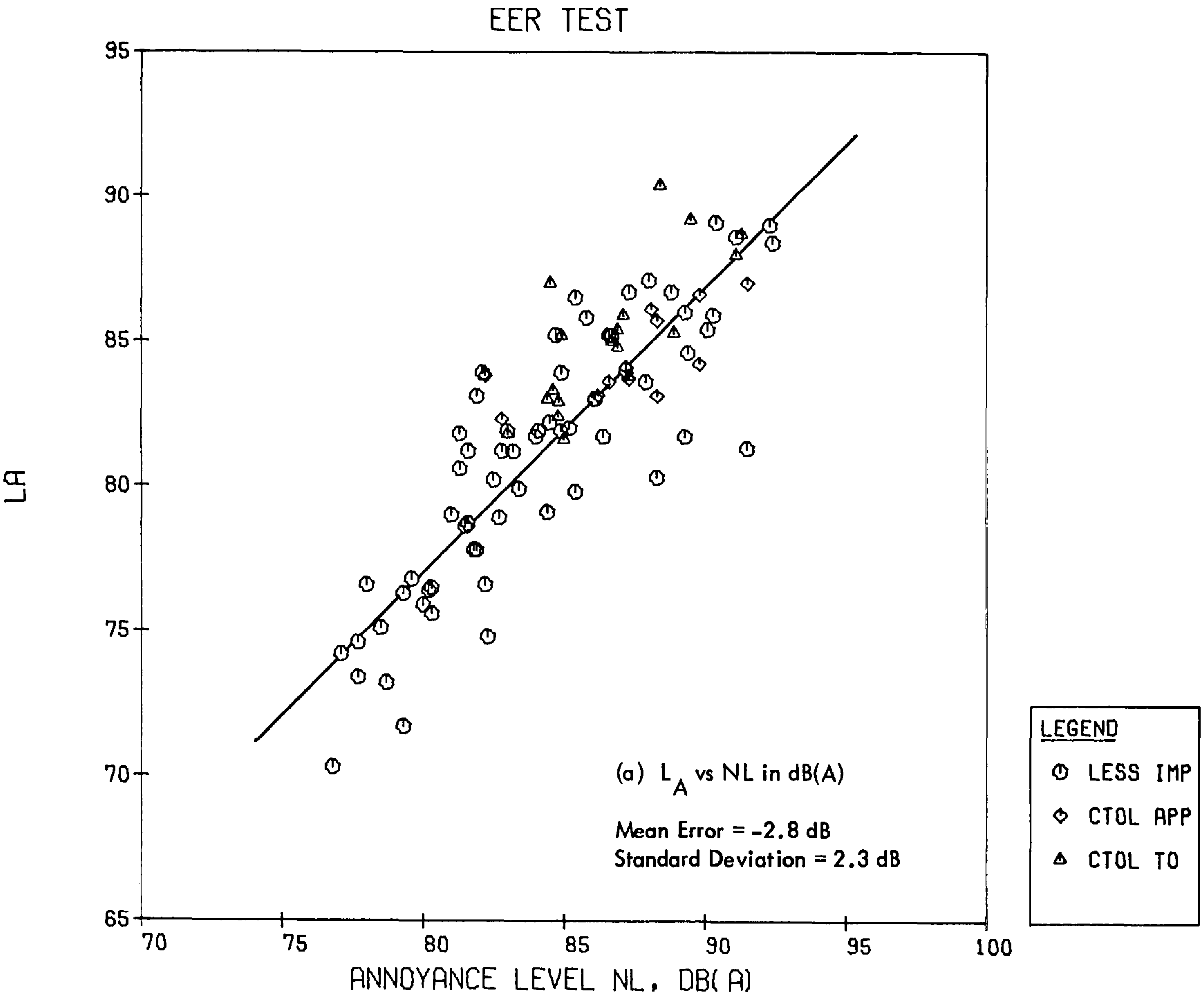


Figure 20. Measured Level Versus Annoyance Level; Exterior Effects Room

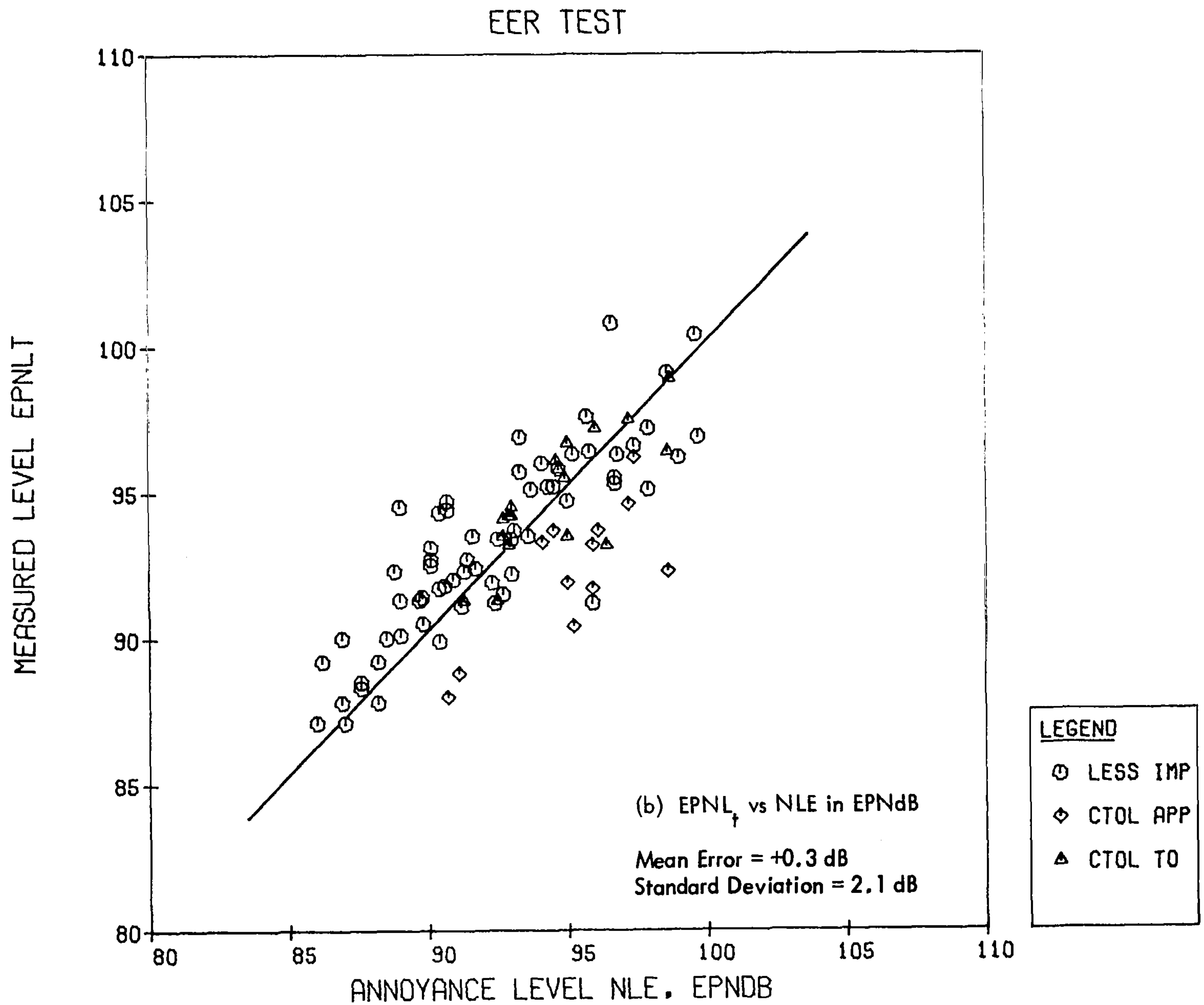


Figure 20. (Continued)

Table 14

High Level Headphone Experiment Annoyance Prediction Errors, dB

Scale	Maximum Levels: All 89 Sounds		Maximum Levels: 59 Helos / 30 CTOLs		Maximum Levels: 45 less imp. / 14 more imp.[2] / 12 approach / 18 takeoff		Time-Integrated Levels: All 89 Sounds		Time-Integrated Levels: 59 Helos / 30 CTOLs		Time-Integrated Levels: 45 less imp. / 14 more im / 12 approach / 18 takeoff	
L_A	-6.3	(2.5)[1]	+0.3	(2.6)	+0.8	(2.2)	-18.6	(2.7)	+1.1	(2.1)	+1.4	(2.0)
					**							
					-1.3	(3.1)					+0.5	(2.5)
									**			
			-0.6	(2.1)	-1.3	(2.1)			-2.4	(2.4)	-4.5	(1.7)
											**	
					-0.1	(1.9)					-0.9	(1.6)
L_D	0.7	(2.3)	+0.2	(2.4)	+0.7	(2.0)	-11.8	(2.3)	+1.1	(1.9)	+1.3	(1.7)
					**	(*)						(*)
					-1.4	(3.0)					+0.6	(2.5)
									**			
			-0.3	(2.1)	+0.2	(2.1)			-2.1	(1.6)	-3.1	(1.6)
											**	
					-0.7	(2.1)					-1.4	(1.2)
L_E	-1.3	(2.4)	+0.5	(2.4)	+0.9	(2.2)	-13.6	(2.8)	+1.3	(2.2)	+1.3	(2.1)
					**							
					-0.7	(2.9)					+1.1	(2.5)
			**						**			
			-0.9	(1.9)	-1.1	(2.1)			-2.6	(2.1)	-4.4	(1.6)
											**	
					-0.8	(1.9)					-1.4	(1.6)
L_F	-0.9	(2.7)	-0.3	(2.8)	+0.3	(2.4)	-13.6	(2.0)	+0.7	(1.9)	+1.0	(1.6)
					**						*	(*)
					-2.3	(3.2)					-0.4	(2.4)
									**	(*)		
			+0.6	(2.3)	+1.6	(2.1)			-1.4	(1.3)	-1.7	(1.5)
					*							
					-0.2	(2.3)					-1.2	(1.1)
PNL	7.1	(2.2)	+0.4	(2.3)	+0.8	(2.0)	-5.0	(2.7)	+1.3	(1.9)	+1.4	(1.8)
					**							
					-1.0	(2.7)					+1.0	(2.3)
			*						**			
			-0.7	(1.9)	-0.8	(1.9)			-2.6	(2.0)	-4.2	(1.7)
											**	
					-0.7	(2.0)					-1.5	(1.3)

(1) Mean errors for subsamples are relative to overall mean error listed for all sounds. Standard Deviation in parentheses throughout table.

(2) More impulsive sounds are those with $EPNL_{ti} - EPNL_t \geq 4$ in Main test.

Table 14 (Continued)

Scale	Maximum Levels: All 89 Sounds	Maximum Levels: 59 Helos / 30 CTOLs	Maximum Levels: 45 less imp. / 14 more imp. / 12 approach / 18 takeoff	Time-Integrated Levels: All 89 Sounds	Time-Integrated Levels: 59 Helos / 30 CTOLs	Time-Integrated Levels: 45 less imp. / 14 more imp. / 12 approach / 18 takeoff
			+0.7 (2.0)			+1.4 (1.7)
		+0.2 (2.5)	** (*)		+1.2 (1.9)	
			-1.3 (3.2)			+0.7 (2.3)
PNL_t	8.9 (2.5)			-3.5 (2.5)	**	
			+0.2 (2.0)			-3.6 (1.7)
		-0.5 (2.4)			-2.4 (1.8)	**
			-0.9 (2.6)			-1.6 (1.4)
			+0.7 (2.4)			+1.2 (2.1)
		+1.0 (2.7)	(*)		+1.8 (2.3)	
			+1.9 (3.5)			+3.6 (2.1)
PNL_{ti}	11.1 (2.9)	**		-1.6 (3.4)	**	
			-1.4 (2.6)			-4.9 (1.9)
		-1.8 (2.5)			-3.6 (1.9)	**
			-2.0 (2.5)			-2.8 (1.4)

Table 15

IER Experiment Annoyance Prediction Errors, dB[1]

	Maximum Levels			Time-Integrated Levels		
Scale	All 93 Sounds	63 Helos / 30 CTOLs	50 less imp. / 13 more imp.[2] / 12 approach / 18 takeoff	All 93 Sounds	63 Helos / 30 CTOLs	50 less imp. / 13 more imp. / 12 approach / 18 takeoff
			+0.1 (2.0)			+0.5 (2.0)
		-0.2 (2.0)	**		+0.5 (1.9)	
			-1.5 (1.4)			+0.4 (1.7)
L_A	-0.9 (1.9)			-13.8 (2.2)	** (*)	
			-0.3 (1.9)			-3.4 (1.7)
		+0.5 (1.8)			-1.0 (2.5)	**
			+1.0 (1.6)			+0.5 (1.6)
			+0.2 (2.1)			+0.5 (2.2)
		0.0 (2.1)			+0.6 (2.1)	
			-0.8 (1.8)			+0.9 (1.7)
L_D	+5.5 (2.0)			-7.0 (2.2)	**	
			-0.6 (2.2)			-2.9 (1.5)
		-0.2 (1.9)			-1.2 (2.0)	**
			+0.7 (1.6)			0.0 (1.4)
			+0.1 (1.8)			+0.5 (2.4)
		-0.1 (1.8)			+0.5 (2.3)	
			-0.7 (1.6)			+0.7 (2.0)
L_F	+2.6 (1.7)			-8.5 (2.2)	** (*)	
			+0.4 (2.3)			-2.4 (1.6)
		+0.3 (1.7)	(*)		-1.1 (1.7)	**
			+0.3 (1.3)			-0.2 (1.2)
			+0.1 (1.9)			+0.5 (2.1)
		-0.1 (1.9)			+0.5 (2.0)	
			-1.0 (1.6)			+0.7 (1.7)
PNL	+12.4 (1.9)			-0.2 (2.2)	**	
			-0.7 (1.8)			-3.0 (1.7)
		+0.1 (1.7)	*		-1.2 (2.1)	**
			+0.7 (1.4)			0.0 (1.4)
			+0.3 (2.1)			+0.7 (2.0)
		+0.0 (2.1)	*		+0.7 (1.9)	
			-1.3 (1.8)			+0.8 (1.7)
PNL_t	+13.8 (2.1)			0.6 (2.3)	**	
			-0.7 (2.4)			-3.4 (1.8)
		+0.1 (2.0)			-1.4 (2.3)	**
			+0.5 (1.7)			-0.1 (1.4)

(1) Mean errors for subsamples are relative to overall mean error listed for all sounds. Standard Deviation in parentheses throughout table.

(2) More impulsive sounds are those with $EPNL_{ti} - EPNL_t \geq 4$ in Main test.

Table 15 (Continued)

Scale	Maximum Levels: All 93 Sounds	Maximum Levels: 63 Helos / 30 CTOLs	Maximum Levels: 50 less imp. / 13 more imp. / 12 approach / 18 takeoff	Time-Integrated Levels: All 93 Sounds	Time-Integrated Levels: 63 Helos / 30 CTOLs	Time-Integrated Levels: 50 less imp. / 13 more imp. / 12 approach / 18 takeoff
			+0.3 (2.1)			+0.7 (2.1)
		0.0 (2.1)	*		+0.7 (2.0)	
			-1.2 (1.9)			+0.8 (1.8)
PNL_{ti}	+14.0 (2.1)			+0.8 (2.3)	**	
			-0.8 (2.3)			-3.6 (1.8)
		0.0 (2.0)			-1.4 (2.4)	**
			+0.5 (1.7)			0.0 (1.5)

Table 16

EER Experiment Annoyance Prediction Errors, dB

Scale	Maximum Levels						Time-Integrated Levels					
	All 93 Sounds		63 Helos / 30 CTOLs		50 less imp. 13 more imp.[1] / 12 approach 18 takeoff		All 93 Sounds		63 Helos / 30 CTOLs		50 less imp. 13 more imp. / 12 approach 18 takeoff	
					-0.0	(2.4)					+0.5	(1.5)
			-0.3	(2.3)					+0.4	(1.6)		
					-1.4	(1.7)					+0.2	(1.9)
L_A	-2.8	(2.3)	*				-13.6	(2.1)	**	(**)		
					-0.1	(2.0)					-3.2	(1.8)
			+0.8	(1.9)	*				-0.8	(2.6)	**	
					+1.3	(1.7)					+0.8	(1.8)
					-0.1	(2.6)					+0.4	(1.8)
			-0.2	(2.5)					+0.5	(1.9)		
					-0.6	(2.1)					+1.0	(2.4)
L_D	+2.7	(2.3)		(*)			-8.0	(2.2)	**			
					0.0	(2.1)					-3.0	(1.6)
			+0.3	(1.8)					-1.1	(2.2)	**	
					+0.5	(1.6)					+0.2	(1.6)
					-0.3	(2.1)					+0.5	(1.9)
			-0.5	(2.1)					+0.5	(1.9)		
					-1.4	(1.7)					+0.7	(2.1)
L_F	0.0	(2.2)	**				-10.2	(1.9)	**			
					+1.4	(2.5)					-2.1	(1.4)
			+1.1	(2.0)					-0.8	(1.7)	**	
					+0.9	(1.6)					0.0	(1.3)
					0.0	(2.3)					+0.5	(1.8)
			-0.1	(2.2)					+0.6	(1.9)		
					-0.5	(1.8)					+1.0	(2.2)
PNL	+9.5	(2.1)					-1.0	(2.2)	**			
					-0.2	(1.8)					-3.2	(1.6)
			+0.1	(1.7)					-1.2	(2.2)	**	
					+0.4	(1.5)					+0.1	(1.5)
					0.1	(2.4)					+0.5	(1.8)
			-0.1	(2.3)					+0.6	(1.9)		
					-0.7	(1.9)					+0.8	(2.2)
PNL_t	+11.0	(2.1)		(*)			+0.3	(2.1)	**			
					0.0	(1.7)					-3.1	(1.6)
			+0.1	(1.7)					-1.3	(2.1)	**	
					+0.2	(1.8)					0.0	(1.4)

Table 16 (Continued)

Scale	Maximum Levels: All 93 Sounds	Maximum Levels: 63 Helos / 30 CTOLs	Maximum Levels: 50 less imp. 13 more imp. / 12 approach 18 takeoff	Time-Integrated Levels: All 93 Sounds	Time-Integrated Levels: 63 Helos / 30 CTOLs	Time-Integrated Levels: 50 less imp. 13 more imp. / 12 approach 18 takeoff
			+0.0 (2.3)			+0.4 (1.8)
		+0.0 (2.2)			+0.6 (1.9)	
			-0.1 (2.0)			+1.2 (2.1)
PNL_{ti}	+11.5 (2.1)			+0.6 (2.2)	**	
			-0.2 (1.7)			-3.2 (1.6)
		-0.1 (1.8)			-1.3 (2.2)	**
			+0.1 (1.8)			0.0 (1.4)

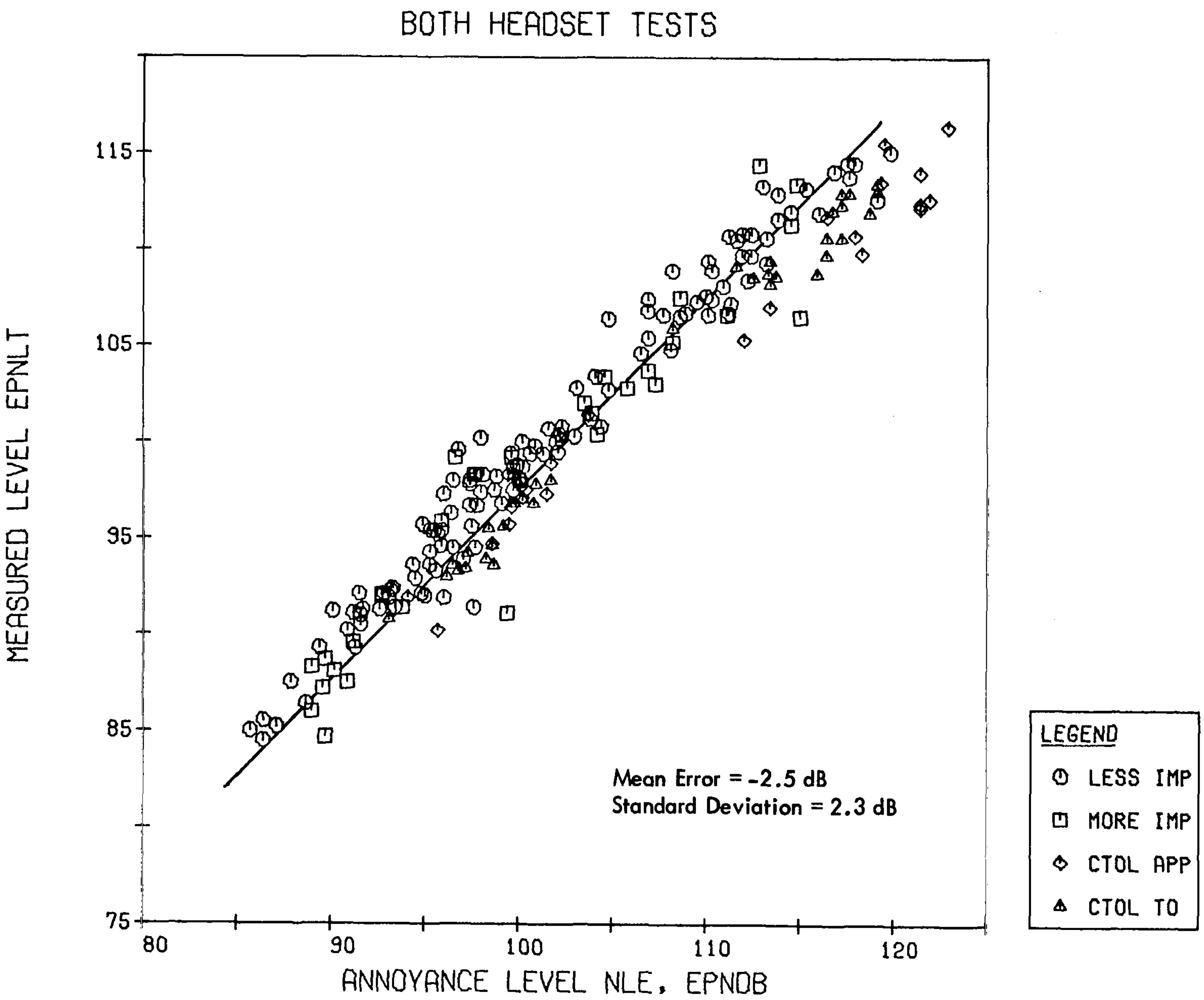


Figure 21. Measured Level Versus Annoyance Level, EPNdB; Combined Headphone Data

There is a distinct divergence between the two samples at the higher levels with the helicopter data retaining a rather more linear relationship. These points are perhaps more evident in Figures 22(a) and 22(b) which show annoyance level plotted against $EPNL_t$ (axes reversed) (a) for the combined helicopter/CTOL sample, and (b) for the helicopters alone.

It will be seen later that the CTOL sounds, particularly the approaches, are characterized by substantial high frequency energy and it can be conjectured that the nonlinearity is associated with the effects of this at the higher test levels. It is noticeable, for example, that the approach sounds diverge more than the takeoff sounds. If this is the explanation, then it points to a deficiency in the frequency-weighting functions, suggesting that as level increases, proportionately more emphasis should be given to the higher frequencies.

Most of the significant differences between the results in Tables 4 and 14 (low and high level headphone tests) are attributable to the divergent results for CTOL approaches. On average, these are underrated by about 2 dB in the high level tests. Otherwise the high level tests corroborate the low level tests quite closely and most of the conclusions outlined in Section 4 are supported. In particular, the duration correction is beneficial (this is best gauged from the subgroup error deviations; the values for the total test sample are increased by the deviant CTOL sounds), the helicopter sounds are overrated relative to the CTOL sounds (by a somewhat greater margin) and the impulse correction, although yielding some small improvement in the consistency of prediction for the more impulsive helicopter sample, does not generally improve the performance of $EPNL_t$ for helicopters.

Turning now to the loudspeaker tests, the IER results in Table 15 show the same (approximately) 2 dB difference between helicopter and CTOL annoyance levels when measured on the time-integrated scales.* However, there is no general improvement associated with the duration correction, probably because the 10 dB-down durations of the low level signals in the IER are significantly affected by background noise, and there is no appreciable difference between the predictive consistency of the different scales. Finally, there are significant differences between the mean prediction errors for CTOL approaches and takeoffs which are very similar to those of the high level headphone tests.

*For both the IER and EER tests, the E-weighted sound levels are not available.

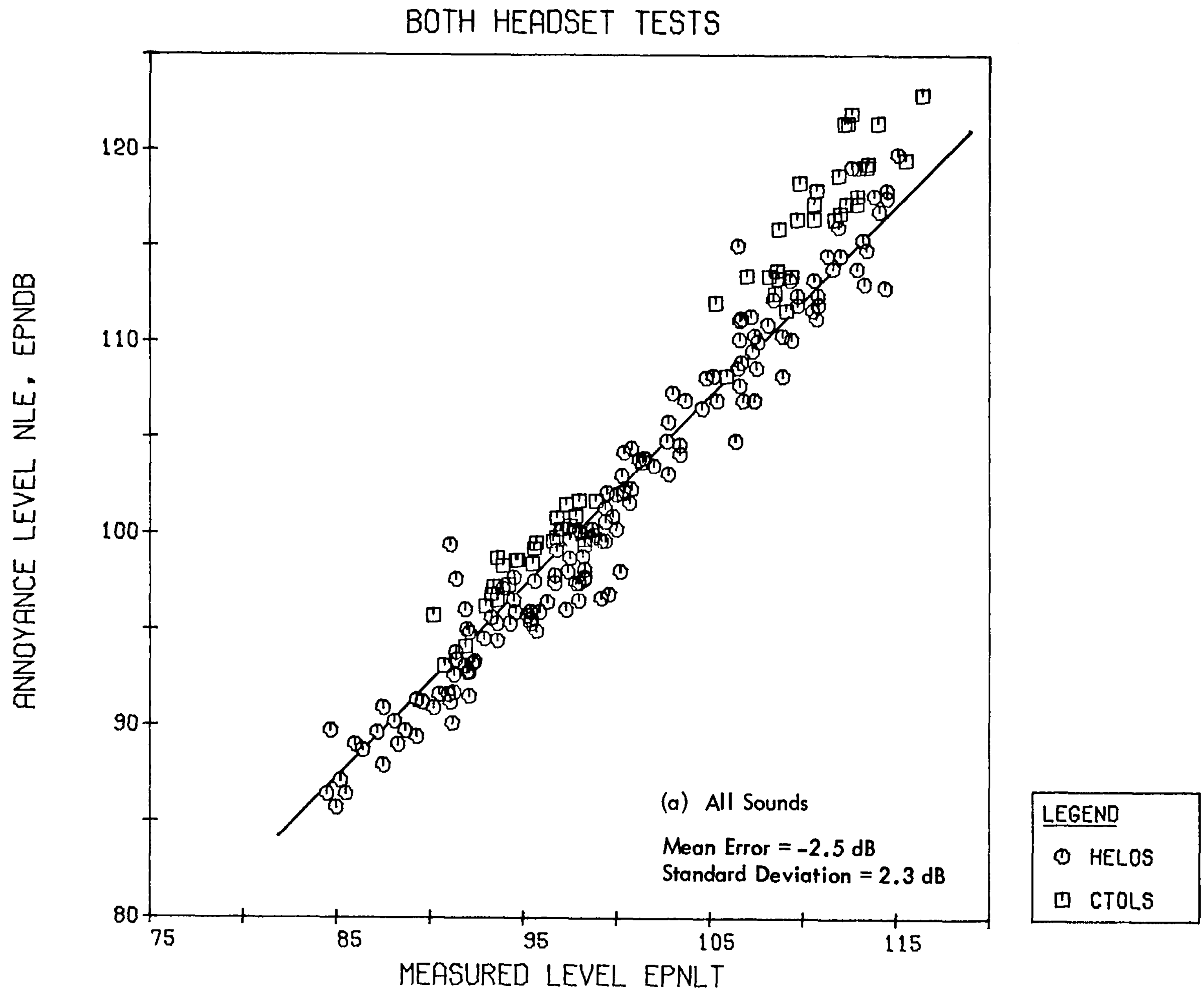


Figure 22. Annoyance Level Versus Measured Level, EPNdB; Combined Headphone Data

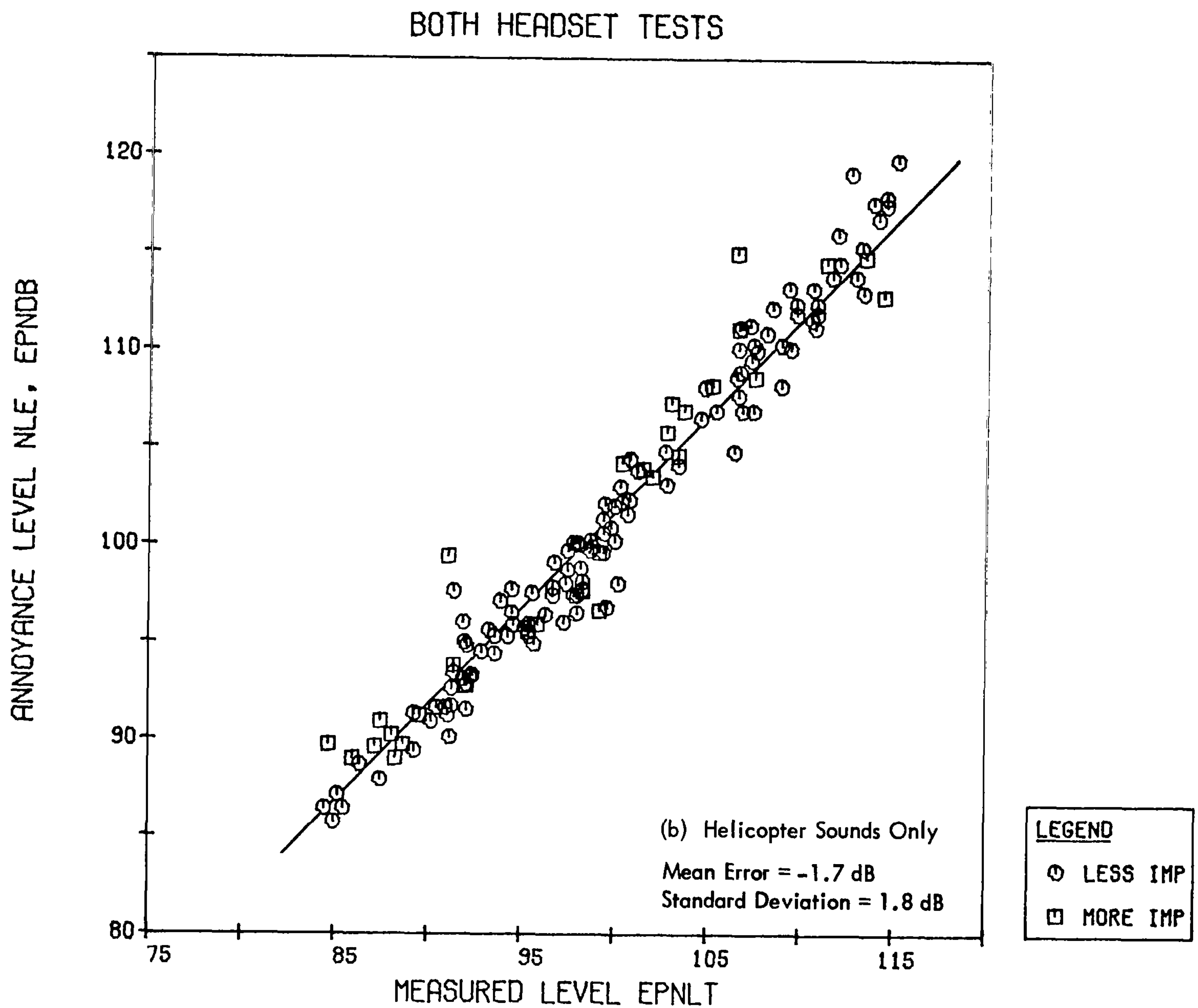


Figure 22 (Continued)

Much the same picture emerges from the EER results summarized in Table 16; indeed, these resemble the IER results quite closely with the important exception that the duration correction reduces the error variance in many instances, if to a slightly lesser extent than was found in the low level headphone tests. In most cases (e.g., for $EPNL_t$), a reduction in scatter for the helicopters is accompanied by an increase in scatter for the CTOLs (the single exception is EL_F). Closer inspection reveals that the increased scatter for the CTOLs is mainly due to a greater than 3 dB mean difference which again arises between the approach and takeoff CTOL groups.

Tables 15 and 16 again distinguish between more and less impulsive helicopters for the IER and EER tests. However, in these tests, the division is an artificial one in that none of the sounds were particularly impulsive as reproduced in the Langley test chambers. In both cases therefore, the "more impulsive" sounds are merely those for which the integrated correction ($EPNL_{ti}$ - $EPNL_t$) was at least equal to 4.0 dB in the headphone tests. It is instructive that, in both loudspeaker tests, there are still no significant differences between the mean prediction errors for the two helicopter samples. Comparing the EER results with those of the main experiment for EPNL and $EPNL_t$, the mean difference has fallen by 1 dB, e.g., relative to the "less impulsive" helicopters; the "more impulsive" ones are on average 1 dB or less annoying when their impulsiveness is removed. This change is barely significant and reinforces the conclusion that the effects of impulsiveness are adequately represented by the conventional, uncorrected noise measurement scales.

The significant differences between CTOL approaches and departure sounds which arose consistently in all three duplicate tests but were not found in the main low level headphone tests pose something of a dilemma. Their presence in the high level headphone tests was tentatively attributed to possible changes in the weighting curves which occur at the very high sound levels. This explanation is clearly not appropriate to the loudspeaker results.

Neither can the IER and EER CTOL results be attributed to different frequency response characteristics of the sound presentation systems. Figure 23 compares typical spectra for CTOL approaches and departures as heard in the headphone and EER tests (these are based on the time-averaged spectra of the approach and takeoff sounds of eight CTOL types). These indicate that in the headphone tests, the measured levels of the CTOL approaches are dominated by

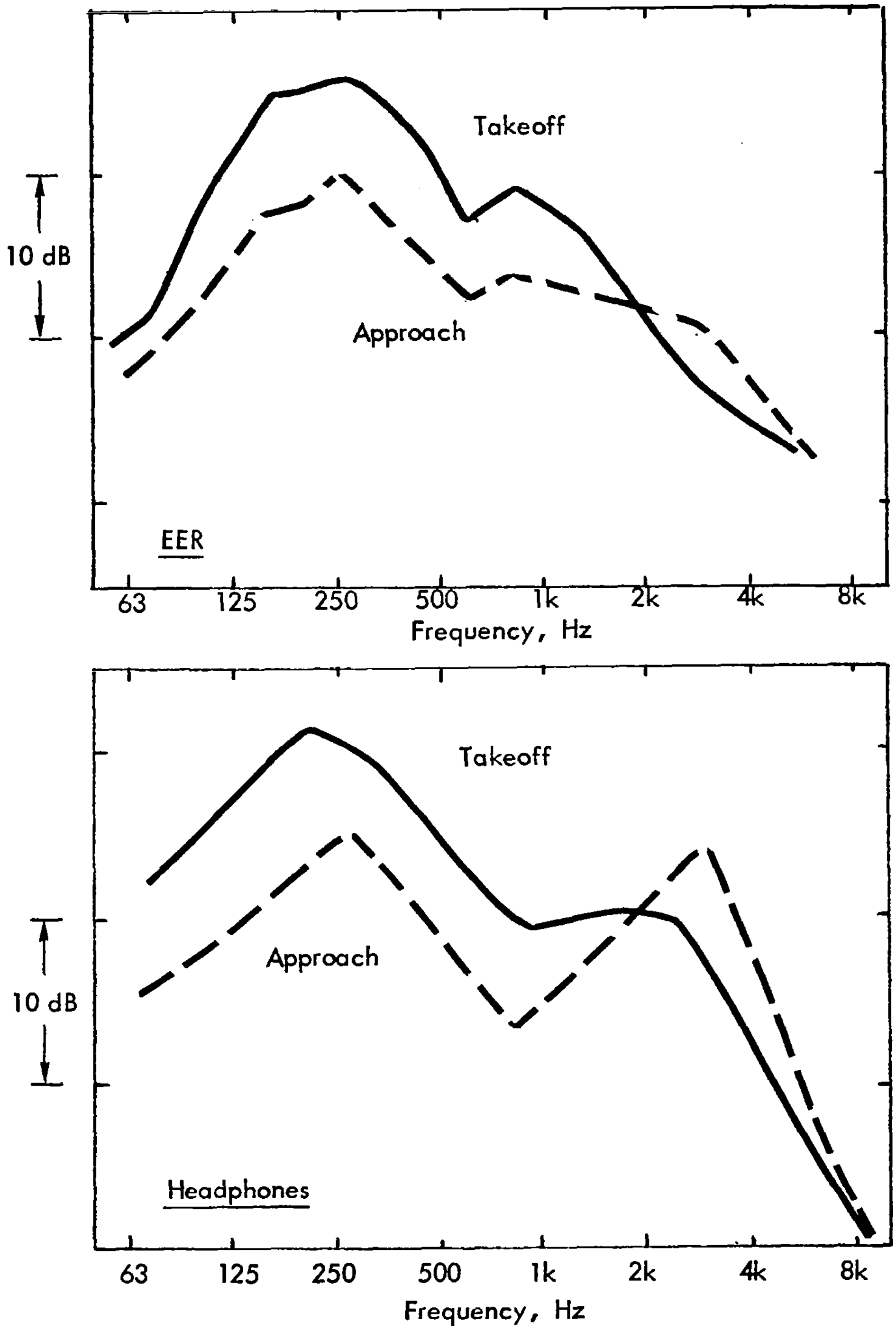


Figure 23. Typical Average Spectra for CTOL Approaches and Takeoffs

the high frequency energy between 2 and 4 kHz. In the EER tests, emphasis shifts to the low frequencies as it does for the takeoff sounds in both tests. Thus, it might be surmised that the 3 dB (average) deficiency in the prediction of annoyance levels for CTOL approaches in the EER tests could be corrected by shifting weight from low frequencies to high in the noise measurement scales. (Note, for example, that proceeding from the A to D to F-weighting, the difference reduces from 4 to 3 to 2 dB.) Attempts have been made to optimize the shape of the weighting function to achieve this end following the technique described in Section 4.4. However, changes which harmonized the EER results caused adverse effects in the headphone results; it has, of course, been noted already that on the basis of the headphone results, the measurement scales are improved for helicopters by a shift of emphasis from high frequencies to low rather than vice versa.

6.0 CONCLUSIONS

Approximately 140 individual helicopter flyover recordings were obtained via the members of ICAO Working Group B. Of these, 89 were of sufficient quality and sufficiently different to include in the study. This was rather less than the 200 or so originally hoped for and it was not possible to achieve the desired degree of independence between the variables of interest (duration, tones, impulsiveness, and frequency distribution). Thirty CTOL recordings, mostly of jet transport aircraft, were included for comparison, particularly to provide a standard of performance for $EPNL_t$* and other noise measurement scales.

The main experiment was performed using headphone presentation to the test subjects and the maximum sound levels of the 119 test sounds covered the range 69 to 93 dB(A). A large part of the experiment was duplicated three times using different subjects and different test conditions.

The test method was based on a rating scale procedure by which each sound was assigned an average annoyance score. This annoyance score was then transformed to an annoyance level defined as the sound level, in decibels, of a common reference sound effectively judged to be equally annoying. The merits of the various noise scaling procedures, including $EPNL_t$, were then assessed in terms of their ability to predict the measured variations in annoyance level between the test sounds.

The main experiment was intended to test the applicability of $EPNL_t$ to as wide a range of helicopter sounds as possible. An original objective of deducing the independent effects of specific underlying variables by multivariate analysis was only achieved to a limited extent due to an unavoidable degree of intercorrelation between the variables.

In the measurement and analysis of the acoustic variables, allowance was made for the frequency response of the test headphones but the impulsiveness correction factors could not be measured directly inside the headphones; instead, they were computed directly from the tape recordings. The true impulse corrections were therefore somewhat less than these nominal values.

The major conclusions drawn from the main experiment were as follows:

---

*The abbreviation $EPNL_t$ is used for the conventional Effective Perceived Noise Level scale used for aircraft noise certification purposes. The subscript t is used explicitly to denote the inclusion of tone corrections since the scale was used with and without these.

1. The Perceived Noise Level scale and the commonly used weighted sound level scales are equivalent in terms of their general ability to predict annoyance level for helicopters, for CTOLs, or for all sounds combined.

2. Conventional duration corrections (+3 dB per doubling of duration) improve the annoyance predicting performance of all the basic scales to which they were applied; duration is a highly significant contributor to judged annoyance.

3. On average, helicopter flyover sounds are judged equally annoying to CTOL sounds when their measured levels are approximately 2 dB higher on the time-integrated scales ($EPNL_t$, EL_A, etc.). In other words, <u>at the same duration corrected levels, helicopters are less annoying than CTOLs.</u>

4. Multiple regression analysis indicated that provided the helicopter/CTOL difference of about 2 dB is taken into account, the particular linear combination of level, duration, and tone corrections inherent in $EPNL_t$ is close to optimum.

5. All scales of time-integrated sound level are very consistent predictors of CTOL noise annoyance levels; for these sounds, the variance of the prediction error is of the same magnitude as that of the estimated experimental error (around 1 dB).

6. All scales of time-integrated level predict the annoyance levels of helicopter noise significantly less consistently than those of CTOL noise. This is probably due to the wide range of acoustic characteristics exhibited by helicopters of different types.

7. The integrated ISO and crest factor impulse correction terms are very highly correlated and may be considered equivalent measures of impulsivity in helicopter noise.

8. <u>Impulse corrections did not improve $EPNL_t$ as a predictor of helicopter noise annoyance.</u> A small but not significant reduction in error variance for the "more impulsive" sounds (defined by a nominal ISO correction of ≥ 4 dB) is more than offset by an increase in variance for "less impulsive" sounds. Furthermore, there is no significant difference between average annoyance levels of the more and less impulsive sounds when equated on any of the time-integrated scales. The impulse correction did not emerge as a significant predictor variable in the multiple regression analysis.

9. The reason that impulse corrections are not effective/not required is attributed to the fact that impulsiveness (a) increases the spectral level of helicopter noise in the frequency range 125-500 Hz, and (b) causes a significant increase in signal duration, which together adequately amplify the sound levels as measured on the conventional scales.

10. Notwithstanding conclusion 1, which is based on the fairly large sample of different helicopter types, there is evidence that the averaging process (over all helicopters) masks significant differences between results for specific helicopter types. Four subgroups of helicopter sounds were classified on the basis of average spectrum shape and a comparison of the mean annoyance prediction errors for these showed clear improvements as emphasis was shifted from high frequencies to low in the sound level weighting functions (A, D, E, and "F"). This may be attributable in part to a correlation between impulsiveness and low frequency content. However, there is a strong likelihood that the conflicting conclusions of previous research into impulsiveness corrections have arisen because of such correlations when attention has been confined to a limited number of helicopter types (especially the Wessex, UH1, and OH58 helicopters).

11. It was found during preliminary experiments that the annoyance judgments of helicopter flyover sounds were unaffected by the long (up to 3 minutes) and very noticeable onset of the sound during the approach of a very impulsive helicopter (Bell 205). This was true even when subjects were specifically instructed to consider signal duration. Accordingly, the "approach component" was not included as a variable in the experiment.

Each of the duplicate experiments involved approximately three-quarters of the test sounds including all the CTOL sounds but only two-thirds of the helicopters. The first was conducted using headphones but with all sound levels nominally 15 dB higher. The second and third were performed simultaneously in the Exterior Effects Room (EER) and Interior Effects Room (IER) at the Langley Research Center using their standard loudspeaker sound replay facilities. All four experiments involved different test subjects.

There were two significant limitations to the Langley loudspeaker tests. In the IER, the signal levels were relatively low (maximum levels between 56 and 73 dB(A)) and the signal-to-background-noise level difference caused significant changes to the duration correction terms. The level range in the EER tests (70-90 dB(A)) was very close to that of the low level headphone tests but in both the IER and the EER, the sound generation systems effectively eliminated impulsiveness from the test sounds.

Taking account of these limitations, the results of all three duplicate experiments broadly agreed with those of the main experiment and thus lend strong support to the generality of the conclusions. In particular, the basic differences in the average annoyance levels of helicopter and CTOL noise was confirmed. Also, the fact that elimination of impulsiveness in the loudspeaker tests did not cause a significant difference to emerge between those subgroups of helicopter sounds which were previously classed as "more" and "less" impulsive, corroborates the conclusion that impulsiveness per se does not contribute more to annoyance than is explained by the increase in level and duration which it causes.

On the negative side, in all three duplicate experiments, the CTOL approach sounds were found to be typically 3 dB more annoying than CTOL takeoff sounds (as measured on the duration-corrected scales). No such difference was found in the main test and this anomaly, for which no plausible explanation can be offered, casts something of a shadow over what is otherwise a surprising consistency between headphone and loudspeaker tests performed with very different groups of over 150 test subjects in different countries.

The results of this study suggest that some previous studies of impulsiveness corrections for helicopter noise indices may have been confounded by interactions between frequency distribution, duration, and impulsiveness. Although this kind of multicolinearity could not be avoided here, the risky consequences of a limited selection of test signals have been minimized. It is concluded that for general prediction of the annoyance-evoking potential of helicopter noise which is not very different in character from that to which we are accustomed, the standard Effective Perceived Noise Level procedure is at least as good as other current noise measurement scales and does not require special provision to penalize impulsiveness. The presence of impulsiveness in a helicopter flyover sound increases both its level and duration to the extent that the increase in the measured time-integrated level accounts for consequent increase in annoyance potential.

This limited endorsement of $EPNL_t$ is not intended to infer that it may be considered an ideal measurement scale for helicopter noise certification. Questions remain concerning the relative contributions of the underlying variables to annoyance and it was found that like other noise scales, $EPNL_t$ is a less consistent predictor of noise annoyance for helicopters than for CTOLs. This is almost certainly due to the considerably wider variety in the various characteristics of helicopter noise which impose a more rigorous test of the noise scaling procedures. This alone points to potential weaknesses in the methodology but other findings reinforce the conclusion that more extensive research into helicopter noise impact is required if a truly equitable noise certification scheme is to be devised. In particular, it is disconcerting that the very long attention-arresting sound of an approaching, highly impulsive helicopter did not affect annoyance judgments in the present experiments. This suggests that in laboratory experiments of this kind, test subjects focus their attention upon the sound of the aircraft as it passes by, perhaps in an attempt to assess its total sound power output. The fact that the sound has a pronounced forward directivity may not influence such judgments. Yet the "hearsay" evidence of complainants near heliports and under helicopter flight routes indicates that the characteristically long audible duration of much helicopter noise is a particular source of aggravation. If this can be established as fact, perhaps by field survey research, the case will be made to develop improved techniques for laboratory study and, ultimately perhaps, to formulate a better concept for helicopter noise certification standards.

REFERENCES

1. International Civil Aviation Organization (ICAO), International Standards and Recommended Practices, "Environmental Protection," Annex 16 to the Convention of International Civil Aviation," First Edition, 1981.

2. "Noise Standards: Aircraft Type Certification," Federal Aviation Regulations Part 36, Federal Register (34 FR 18364), November 1969 through Amendment 12, August 1, 1981.

3. ICAO Committee on Aircraft Noise, Report Document 9171 (CAN 5), Fifth Meeting, Montreal 13-15 November 1976.

4. Munch, C. L. and King, R. J. "Community Acceptance of Helicopter Noise: Criteria and Application." NASA CR-132430, 1974.

5. Magliozzi, B., Metzger, F. B., Bausch, W., and King, R. J., "A Comprehensive Review of Helicopter Noise Literature," Report No. FAA-RD-75-79, June 1975.

6. Williams, R. and Leverton, J. W., "Review of Literature Concerning Subjective Studies of Impulsive Helicopter Noise," Note AA 1315, Westland Helicopters Ltd., 1979.

7. Molino, J. A., "Should Helicopter Noise Be Measured Differently from Other Aircraft Noise? – A Review of the Psychoacoustic Literature." NASA CR-3609, 1982.

8. Berry, B. F., Fuller, H. C., John, A. J., and Robinson, D. W., "The Rating of Helicopter Noise: Development of a Proposed Impulse Correction," NPL Acoustics Report AC 93, December 1979.

9. Powell, C. A., "Subjective Field Study of Response to Impulsive Helicopter Noise," NASA Technical Paper 1833, April 1981.

10. Shepherd, K. P., "A Laboratory Study of the Subjective Response to Helicopter Blade-Slap Noise," NASA CR-158973, 1978.

11. ICAO Committee on Aircraft Noise, Report Document 9286 (CAN 6), Sixth Meeting, Montreal, 23 May to 7 June 1979.

12. Federal Aviation Administration, Proposed Rule, "Noise Standards for Helicopters in the Normal, Transport, and Restricted Categories," NPRM 44 FR 42410, July 19, 1979 (subsequently withdrawn).

13. Ollerhead, J. B., "An Evaluation of Methods for Scaling Aircraft Noise Perception," NASA CR-1883, October 1971.

REFERENCES (Continued)

14. Stevens, S. S., "Perceived Level of Noise by Mark VIII and Decibels (E)," J. Acoust. Soc. Am. 51, 575-593, 1972.

15. Powell, C. A. and McCurdy, D. A., "Effects of Repetition Rate and Impulsiveness of Simulated Helicopter Rotor Noise on Annoyance," NASA Technical Paper 1969, February 1982.

APPENDIX A

Subjects' Instructions

These tests are part of an investigation into the characteristics of aircraft noise which cause annoyance to people who live near airports. We would like you to judge how ANNOYING some aircraft and helicopter sounds are.

Through your headphones, you will hear recordings of various aircraft and helicopter sounds. The number of each sound will be announced before it begins. On your score sheet, you will find scales like the one below which you will use to record your judgment of each sound

Not Annoying At All	0	1	2	3	4	5	6	7	8	9	10	Extremely Annoying

After each sound there will be a break of a few seconds. During this interval, please indicate how annoying you consider the sound to be by placing a mark across the scale. If you judge a sound to be only slightly annoying, then place your mark closer to the NOT AT ALL ANNOYING end of the scale. On the other hand, if you judge a sound to be very annoying, then place your mark closer to the EXTREMELY ANNOYING end of the scale. A mark may be placed anywhere along the scale, not just at the numbered locations.

When making your judgment of each sound, consider how you would feel if you heard it at home on a number of occasions during the day and take into account all the characteristics of the sound. THERE ARE NO RIGHT OR WRONG ANSWERS; we are only seeking your personal opinions.

RATING SHEET

Subject No. ________ Group ________ Session ________ Tape ________

Sound

Sound													
1	Not Annoying At All	0	1	2	3	4	5	6	7	8	9	10	Extremely Annoying
2	Not Annoying At All	0	1	2	3	4	5	6	7	8	9	10	Extremely Annoying
3	Not Annoying At All	0	1	2	3	4	5	6	7	8	9	10	Extremely Annoying
4	Not Annoying At All	0	1	2	3	4	5	6	7	8	9	10	Extremely Annoying
5	Not Annoying At All	0	1	2	3	4	5	6	7	8	9	10	Extremely Annoying
6	Not Annoying At All	0	1	2	3	4	5	6	7	8	9	10	Extremely Annoying
7	Not Annoying At All	0	1	2	3	4	5	6	7	8	9	10	Extremely Annoying
8	Not Annoying At All	0	1	2	3	4	5	6	7	8	9	10	Extremely Annoying
9	Not Annoying At All	0	1	2	3	4	5	6	7	8	9	10	Extremely Annoying
10	Not Annoying At All	0	1	2	3	4	5	6	7	8	9	10	Extremely Annoying
11	Not Annoying At All	0	1	2	3	4	5	6	7	8	9	10	Extremely Annoying

APPENDIX B

Summary of Test Sounds

Table B-1 identifies the aircraft flyover sounds which were included in the subjective experiments. Most of the original sound recordings were furnished by members of ICAO Working Group B. These were copied, in whole or in part, onto the various test tapes which were subsequently replayed to the subjects. The acoustic variables listed in the table were measured from the test tapes and the levels correspond to those heard by the subjects in the main experiment. It should be noted that these are not related to the actual flight levels which occurred when the original recordings were made. The recording levels were selected to provide the best possible reproduction quality having regard to both the dynamic range of the signal, the background noise level, and the tape noise level.

For each test sound, the following variables are listed (where known):

1. Aircraft type
2. Recording number
3. Flight mode: Approach (A), Takeoff (T), or Level flyover (L)
4. Height, (m): The nominal altitude of the aircraft at its closest point to the microphone
5. Sideline distance, S (m): The nominal lateral separation between the microphone and the aircraft flight track
6. Speed, V: Usually indicated airspeed (kt)
7. Maximum replay A-weighted sound level, L_A (dB(A))
8. Duration corrected A-weighted sound level, EL_A (dB(A))
9. Total signal duration (approximate) T, sec.
10. Overall ISO Impulse correction $I = EPNL_{ti} - EPNL_t$, (dB)
11. NL, Judged Annoyance level, (dB(A)).

Table B-1

Characteristics of Test Sounds

Manufacturer/Model	Record No.	Flight Mode	Height m	S m	V kt	L_A dB(A)	EL_A dB(A)	T sec	I dB	NL dB(A)
Sikorsky S61	1	A	100	0	70	79.9	76.6	26	3.9	82.7
	2	L	150	0	100	83.3	78.1	17	0.7	86.7
	3	L	150	150	100	85.6	81.9	17	1.2	88.5
Sikorsky S64	1	A	100	0	60	77.0	75.9	22	2.9	82.4
	2	L	150	0	60	76.9	74.4	21	2.5	80.5
	3	L	150	0	105	75.2	70.3	14	0.9	77.6
	4	L	150	0	95	84.0	79.4	29	0.9	87.3
Sikorsky S76	2	T				88.3	85.3	27	2.1	88.1
	3	T				87.1	83.7	33	2.0	89.5
	4	L			140	86.8	82.4	33	1.6	91.0
	5	A				87.1	84.0	32	1.1	92.6
	6	L	150	0	140	87.3	82.6	24	1.6	90.2
Sikorsky CH53	1	A	100	0	91	84.7	81.1	15	1.8	88.6
	2	L	150	0	150	81.6	77.5	13	1.3	85.9
	3	L	150	150	150	86.2	82.3	20	1.0	89.2
Bell 204B	1	L	270	370	110	70.5	69.7	33	2.6	78.2
	2	L	90	0	110	74.9	72.4	22	5.2	81.9
	3	L	90	120	110	68.7	69.1	19	3.3	77.6
	4	L	90	0	110	71.0	69.7	31	3.5	76.9
	5	L	270	0	110	75.9	75.9	40	4.0	84.1
	6	L	90	120	110	74.8	72.7	27	4.3	81.1
	7	L	90	120	110	78.1	76.3	33	3.0	84.5
	8	L	270	370	110	75.5	76.5	54	3.2	84.7
	9	L	270	0	110	77.8	76.4	43	3.9	84.6
	10	L	90	120	110	74.8	74.7	25	3.3	83.0
Bell 205	1	L	150	0	120	68.4	69.3	23	5.6	80.9
	2	L	150	0	60	73.8	72.1	26	4.5	80.7
	3	L	150	0	40	78.1	77.6	33	5.4	83.9
	4	L	150	0	60	74.4	74.6	33	4.6	82.4
	5	L	150	0	120	76.7	73.1	18	4.8	81.3
Bell UH1B	1	L	-	-	120	78.8	77.0	19	5.6	90.5
Bell 206L	1	A	100	0	52	83.2	82.1	29	2.3	87.5
	2	L	150	0	118	82.4	78.3	17	1.4	85.5
	3	L	150	150	118	84.5	82.7	24	1.2	91.3

Table B-1 (Continued)

Manufacturer/Model	Record No.	Flight Mode	Height m	S m	V kt	L_A dB(A)	EL_A dB(A)	T sec	I dB	NL dB(A)
Bell OH58A	1	L	270	0	110	78.4	76.8	26	1.7	84.3
	2	L	90	120	110	81.8	76.6	22	2.2	88.5
	3	L	90	120	110	75.0	72.9	15	1.5	79.1
Bell 212	1	A	100	0	59	74.3	73.7	27	5.4	80.1
	2	L	150	0	107	74.7	73.5	20	3.7	82.3
	3	L	150	150	107	71.2	70.8	20	3.9	79.7
	4	L	150	0	60	82.5	84.7	47	3.9	92.2
Bell 47G	1	A	100	0	52	77.5	76.3	24	4.7	84.9
	2	L	150	0	52	83.0	81.6	17	0.5	88.5
	3	L	150	0	59	84.9	82.9	24	0.8	91.3
Boeing CH47	1	A	150	0	60	79.1	77.6	28	3.2	83.9
	2	L	150	0	60	81.4	80.9	25	4.3	87.0
	3	L	150	0	150	73.3	70.3	15	4.6	80.1
Hughes 500C	1	A	100	0	62	83.6	80.6	20	3.7	87.6
	2	L	150	0	118	86.3	82.3	18	0.6	90.2
	3	L	150	0	118	87.2	84.1	23	1.6	91.1
Westland Wessex	1					93.2	86.9	38	1.7	91.5
	2					88.0	84.0	31	1.6	87.4
	3					88.2	84.5	31	1.8	89.1
	4					85.6	82.9	32	1.7	86.9
	5					86.8	84.0	20	1.8	87.9
	6					90.9	89.1	32	0.9	95.5
Westland Lynx	1	L	150		70	78.6	79.5	33	2.6	88.0
	2	L	150		130	78.8	77.4	37	2.0	82.8
Westland Scout	1	L			90	78.8	76.3	24	1.5	87.0
Westland Sea King	1	L				79.1	78.1	21	0.5	86.1
Aerospatiale Ecureuil (Squirrel)	1	L	300	0	100	80.7	80.2	32	0.6	86.8
	2	L	300	0	110	82.8	80.5	34	1.1	86.4
	3	L	300	0	120	78.7	76.8	30	1.0	82.8
	4	L	150	0	100	88.8	84.6	28	0.4	93.6
	5	L	150	0	110	88.9	84.2	31	0.8	93.4

Table B-1 (Continued)

Manufacturer/Model	Record No.	Flight Mode	Height m	S m	V kt	L_A dB(A)	EL_A dB(A)	T sec	I dB	NL dB(A)
Aerospatiale Gazelle	1	A	100	0	73	82.7	79.8	23	2.7	86.2
	2	L	150	0	130	86.7	82.4	16	0.8	89.6
	3	L	150	150	130	88.9	84.2	15	0.6	91.3
	4	L	300	0	111	87.9	85.2	20	0.5	92.9
	5	L	300	0	125	89.4	86.9	35	0.4	96.1
Aerospatiale Super Frelon	1	L	300	0	105	84.2	80.6	32	0.5	86.9
	2	L	300	0	115	83.7	81.2	28	0.6	85.9
	3	L	300	0	125	81.4	78.4	28	0.7	85.4
	4	L	150	0	115	88.6	85.1	26	0.3	94.3
	5	L	150	0	105	87.6	83.7	24	0.8	88.7
Aerospatiale Puma	1	A	100	0	68	84.6	83.6	19	4.1	88.5
	2	L	150	0	112	87.9	84.3	22	1.5	90.5
	3	L	300	0	134	79.2	76.4	24	0.9	83.6
	4	L	150	0	112	87.5	84.0	22	1.2	92.9
	5	L	300	0	116	86.3	84.8	48	0.7	91.9
	6	L	300	0	127	86.9	84.1	24	0.6	91.1
Boelkow Bo.105	1	A	100	0	68	88.7	85.6	20	4.5	90.5
	2	L	150	0	119	82.9	79.4	23	1.6	88.5
	3	L	150	150	108	81.7	79.9	24	1.7	86.4
	4	L	150	150	120	79.7	77.6	22	2.4	81.3
	5	L	150	150	120	81.5	75.5	18	0.9	82.0
	6	L	150	150	119	86.9	84.0	23	1.9	90.8
	7	L	150	0	119	88.6	85.4	22	4.5	88.0
	8	L	150	150	120	86.1	84.2	27	2.2	89.3
Boeing 707	1	A				92.9	84.4	12	0.3	94.9
	2	T				85.4	80.8	23	0.7	91.1
Boeing 727	1	A				83.9	80.7	19	0.7	90.7
	2	A				84.8	82.2	18	0.2	90.5
	3	T				86.6	79.6	27	1.0	87.3
Boeing 737	1	A				90.8	83.0	19	0.2	92.8
	2	T				83.3	80.7	23	0.7	88.4
Boeing 747	1	A				89.0	83.7	13	0.2	92.6
	2	T				87.1	80.7	15	0.2	92.8
McDonnell Douglas DC-9	1	A				90.2	83.7	21	0.9	93.0
	2	T				83.6	80.7	23	0.7	89.5

Table B-1 (Continued)

Manufacturer/Model	Record No.	Flight Mode	Height m	S m	V kt	L_A dB(A)	EL_A dB(A)	T sec	I dB	NL dB(A)
BAc Trident 2	1	A				86.4	80.3	14	0.3	90.4
	2	A				82.3	75.1	12	0.3	86.8
	3	T				86.7	82.9	29	0.9	91.6
	4	T				85.2	83.2	31	1.3	90.9
BAc Trident 3	1	A				85.0	76.6	10	2.2	85.1
	2	T				89.9	84.4	22	1.1	92.1
BAc 111	1	A				86.2	81.0	17	0.6	91.3
	2	T				84.3	81.7	20	0.8	89.2
Lockheed L1011	1	A				85.5	78.9	11	0.6	89.5
	2	T				88.5	81.0	17	0	90.1
BAc HS125	1	T				86.4	85.3	20	0.8	90.6
Fokker F28	1	T				88.5	84.0	24	0.7	91.1
Aerospatiale Caravelle	1	A				88.6	81.4	17	0.5	91.2
BAc VC10	1	T				82.2	79.6	28	1.0	87.6
	2	T				83.8	79.8	27	0.8	89.8
Airbus Industrie A300	1	T				81.6	79.0	29	0.3	88.3
	2	T				80.7	79.5	19	0.4	89.4
BAc Viscount	1	T				80.2	75.1	16	0.4	84.2
	2	T				83.6	77.8	20	0.3	87.9

APPENDIX C

Representative Time Histories and Spectra of Helicopter Signals

For each of 16 sounds selected to cover a reasonable range of helicopter types and acoustic characteristics, an analysis was made of simultaneous 2-second samples from the tape output and the output from a microphone in a flat plate coupler under one of the headphones. The sampling rate was 5 kHz/channel with the anti-aliasing filter set at 2 kHz. For each sample, a pressure time history covering a few main rotor blade passing intervals over a 0.2 second period or a 0.4 second period is plotted in Figures C-1 and C2, respectively. (Note that ordinal scales in these figures are arbitrary – they are not comparable between samples.) A power spectrum, also using an arbitrary ordinate scale and based on the complete 2-second record, was computed and is shown as the lowest plot in each part of Figures C-1 and C-2. The frequency range for the latter was varied in order to convey the most useful information. The rotor blade passing frequency generally falls in the frequency range of about 12 to 30 Hz and can be most reliably determined from the spacing between the peaks of these spectrum plots which cover a substantial portion of the rotor noise harmonics.

Each sample was taken during the helicopter approach, usually well before the overhead position, when the sound was subjectively dominated by rotor noise. The starting times of the samples, listed in Table C-1, were measured from the approximate start of the recorded signal. Also listed in Table C-1 are the ISO-impulse corrections in decibels for the sample records. In each case, the value for the tape output is given and, when available, the value measured from the headphone/coupler output. Some of the latter values may differ slightly from those specified for the tape output since they were not necessarily made at the same one-half second intervals. Furthermore, these ISO corrections may differ from those given in Table B-1 of Appendix B due to the fact that the latter values were based on analysis runs from the complete record and are not necessarily correlated with the impulsiveness of the short samples presented in this Appendix.

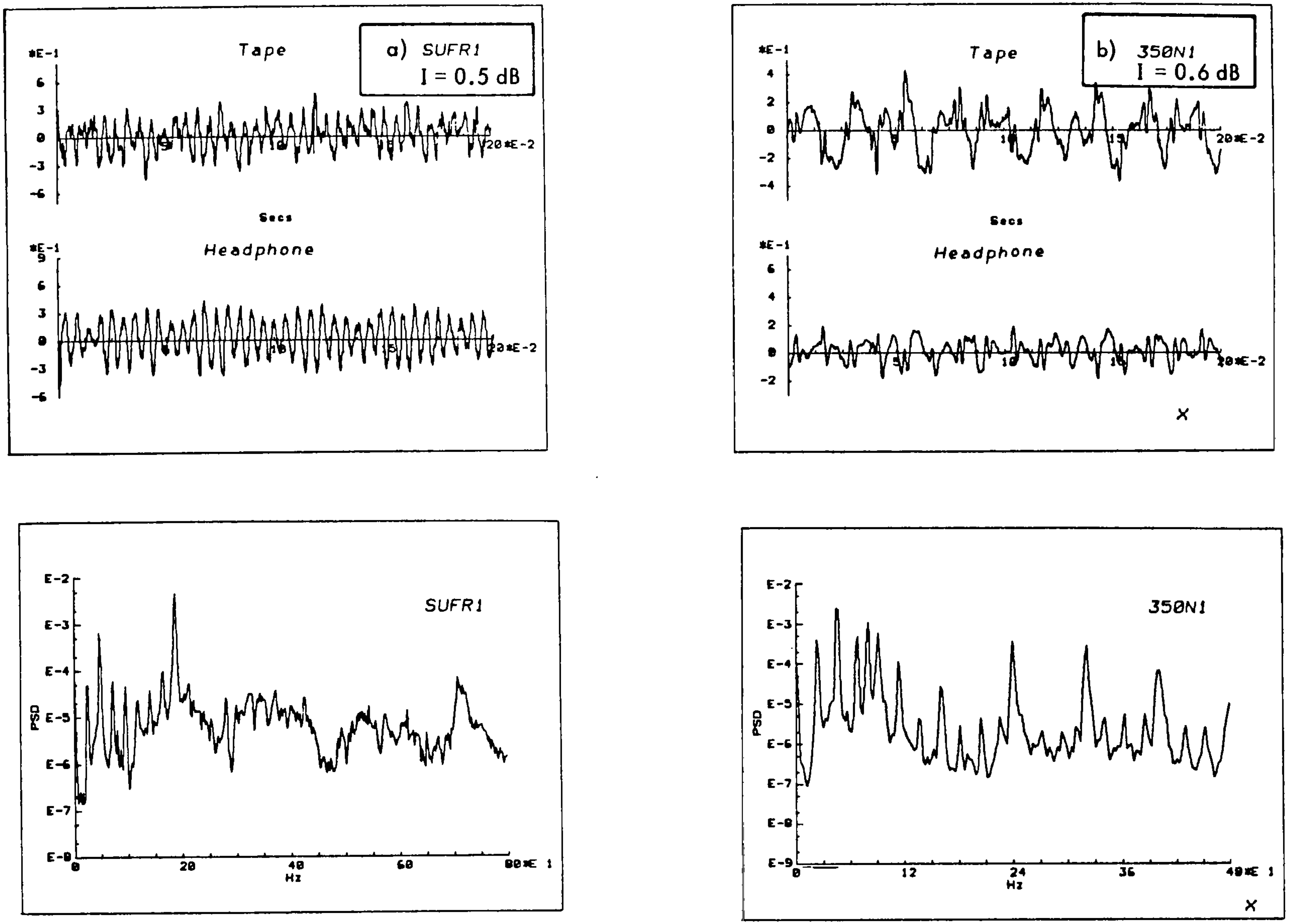


Figure C-1. Time History of 0.2 Second Portion of Signal at Output of Tape (Upper Plot) and Headphone Output in a 6 cc Coupler (Middle Plot) and Narrow Band Spectra of 2 Second Samples (Lower Plot) of Recorded Helicopter Signals - Arranged in Ascending Order of ISO Impulse Correction Factor (I) Based on the 2-Second Sample. (Ordinate scales are arbitrary)

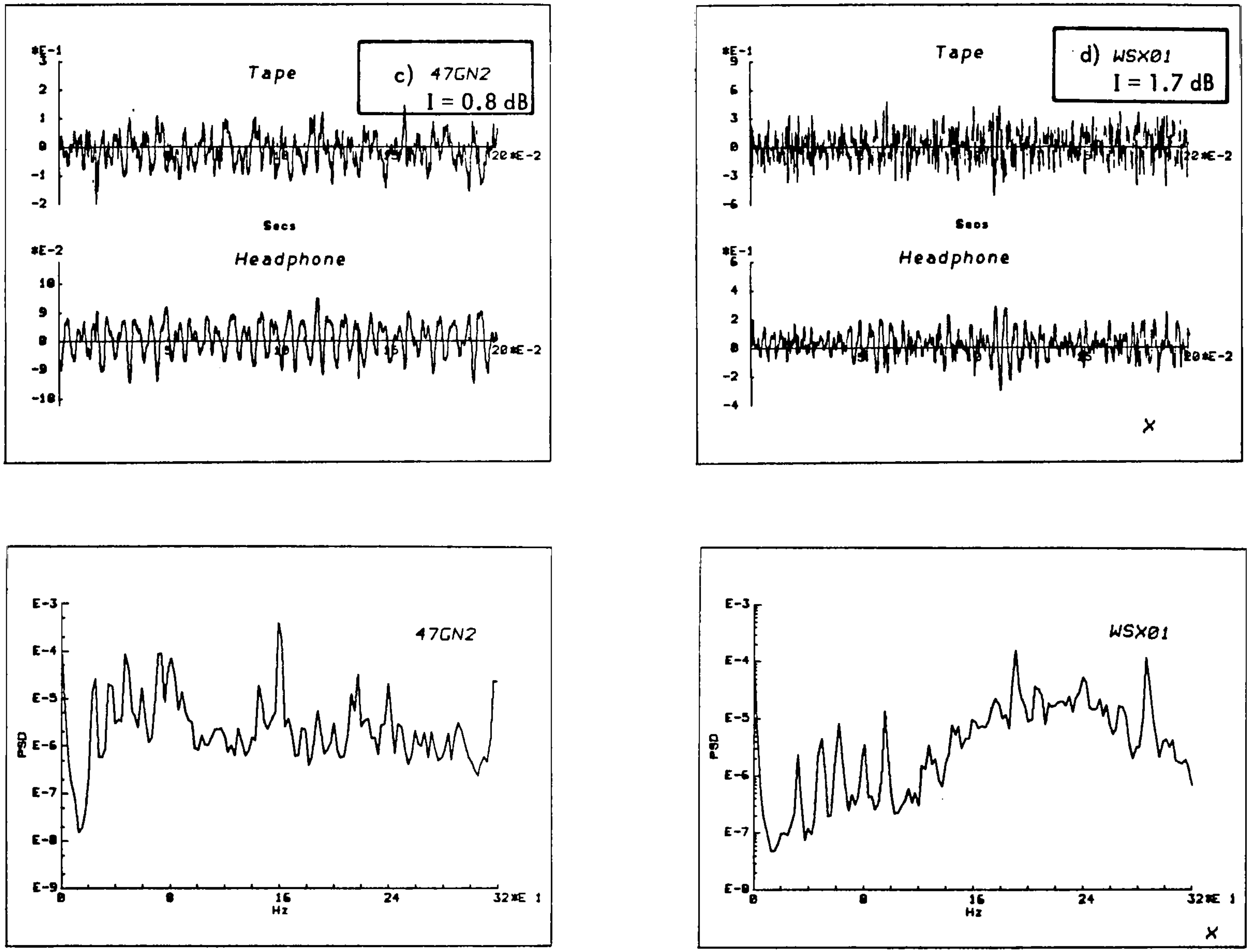
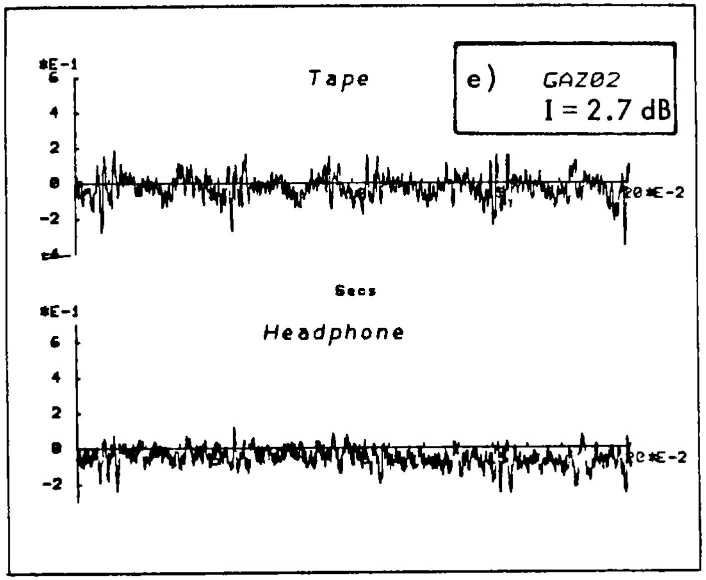
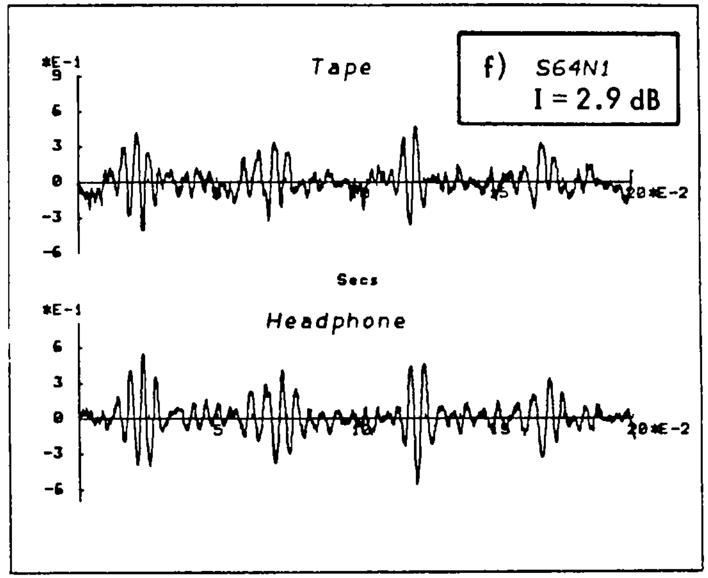


Figure C-1 (Continued)





(spectrum plot not available)

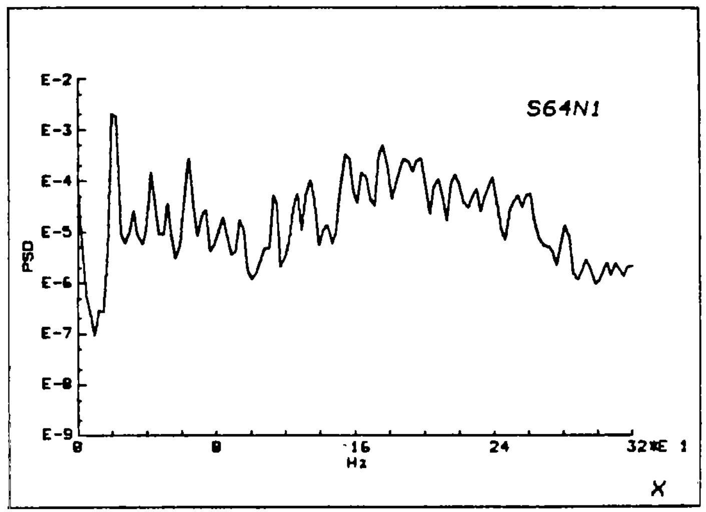


Figure C-1 (Continued)

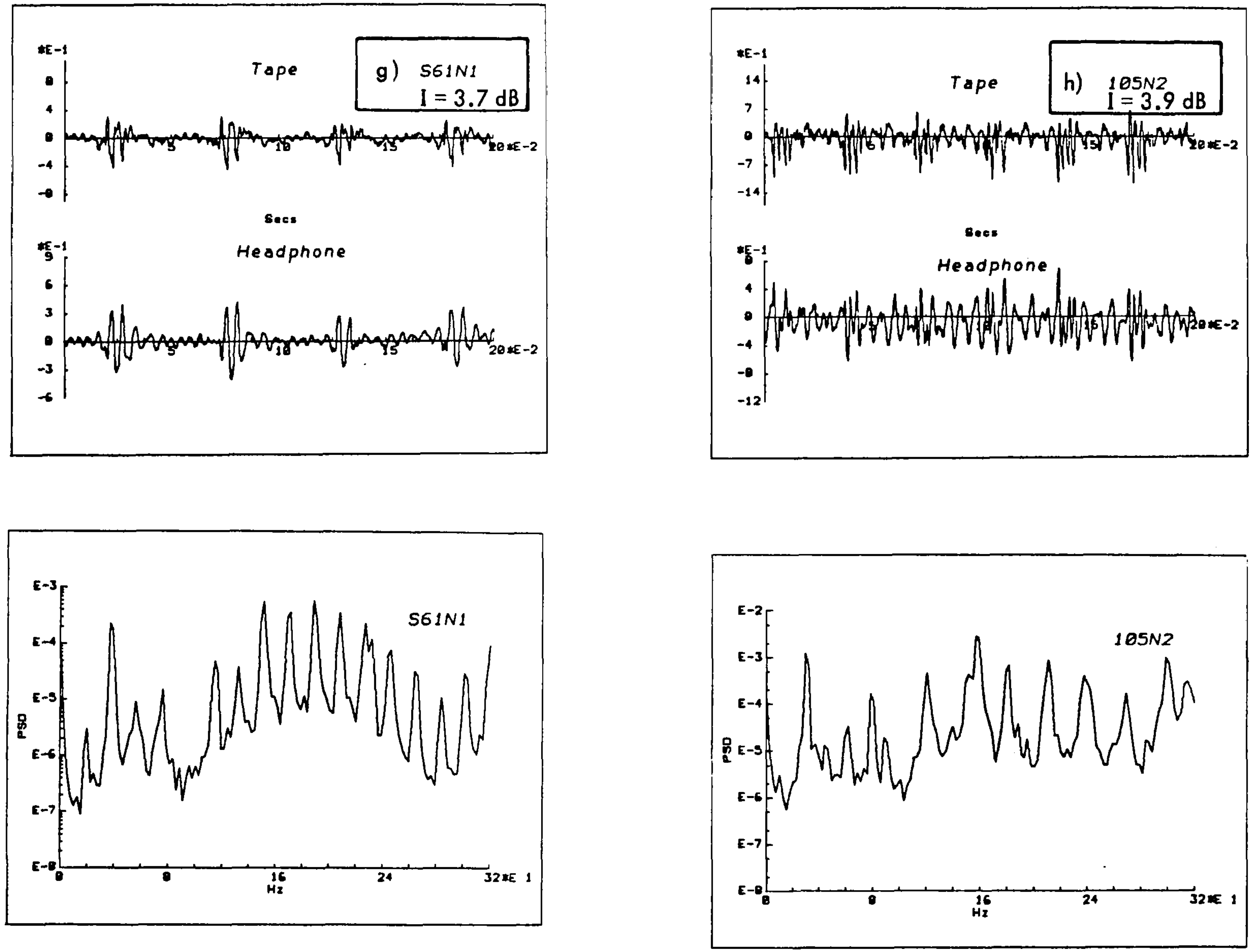


Figure C-1 (Continued)

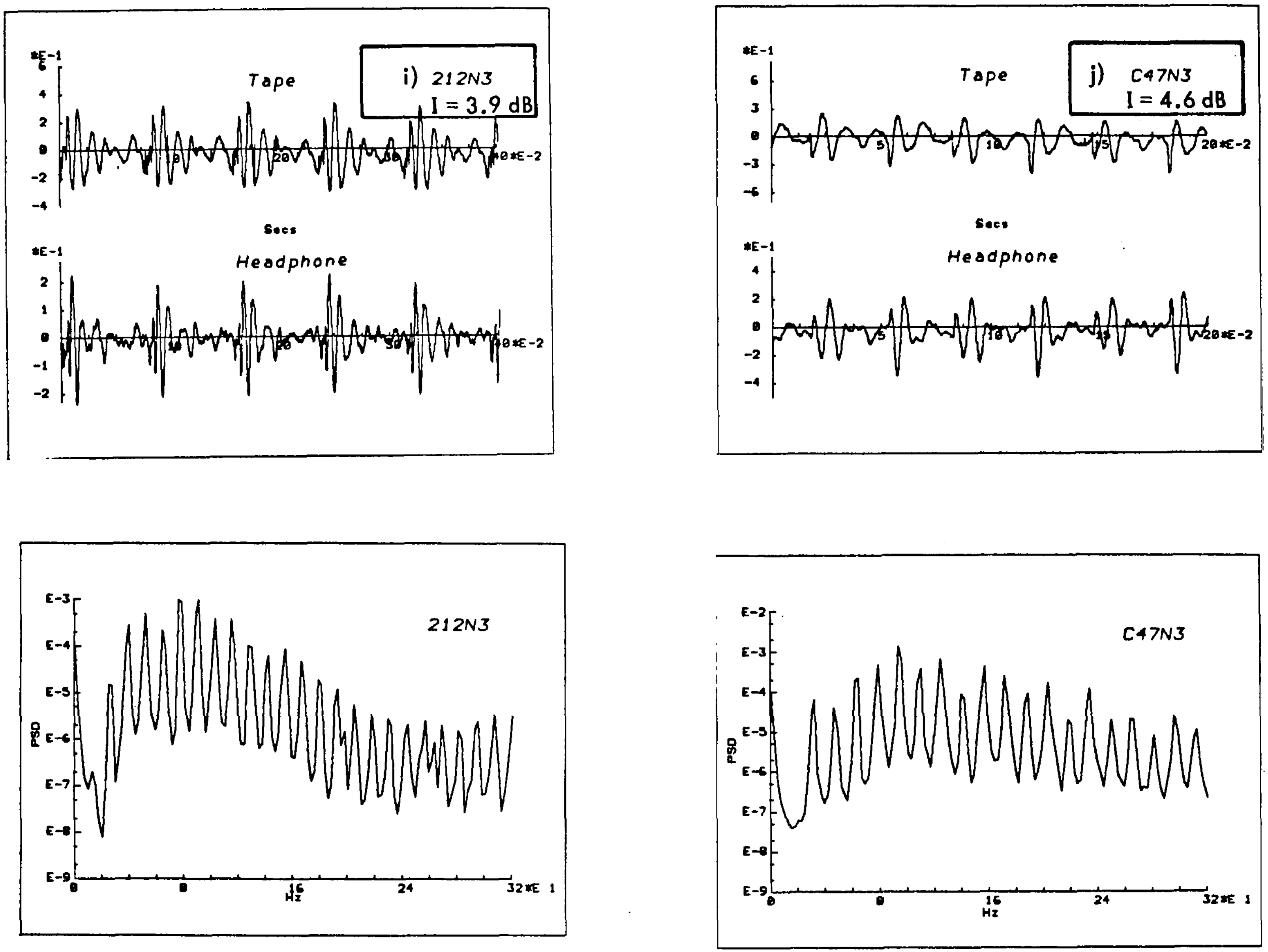


Figure C-1 (Continued)

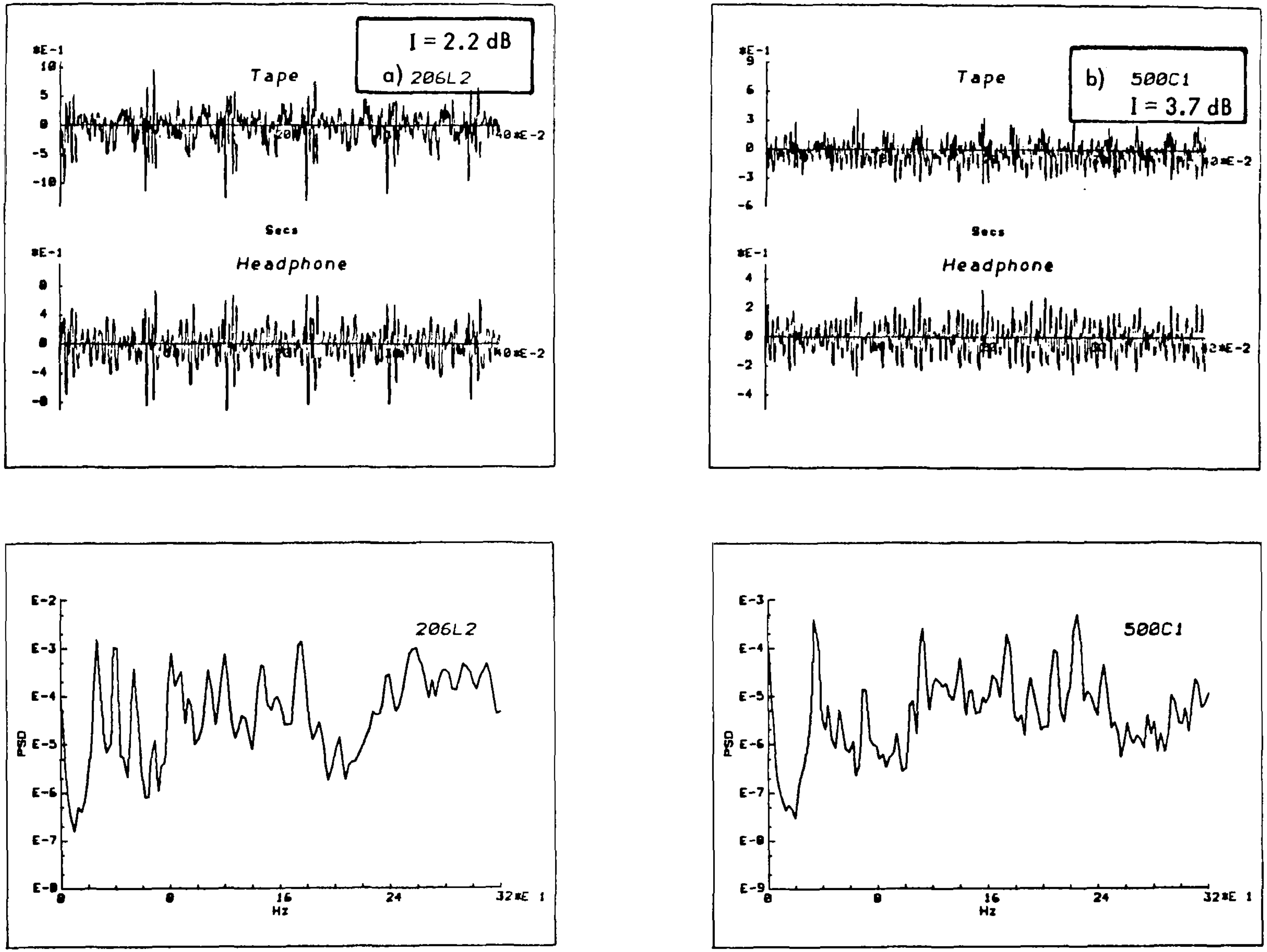


Figure C-2. Time History of 0.4 Second Portion of Signal at Output of Tape (Upper Plot) and Headphone Output in a 6 cc Coupler (Middle Plot) and Narrow Band Spectra of 2 Second Samples (Lower Plot) of Recorded Helicopter Signals - Arranged in Ascending Order of ISO Impulse Correction Factor (I) Based on the 2-Second Sample. (Ordinate scales are arbitrary)

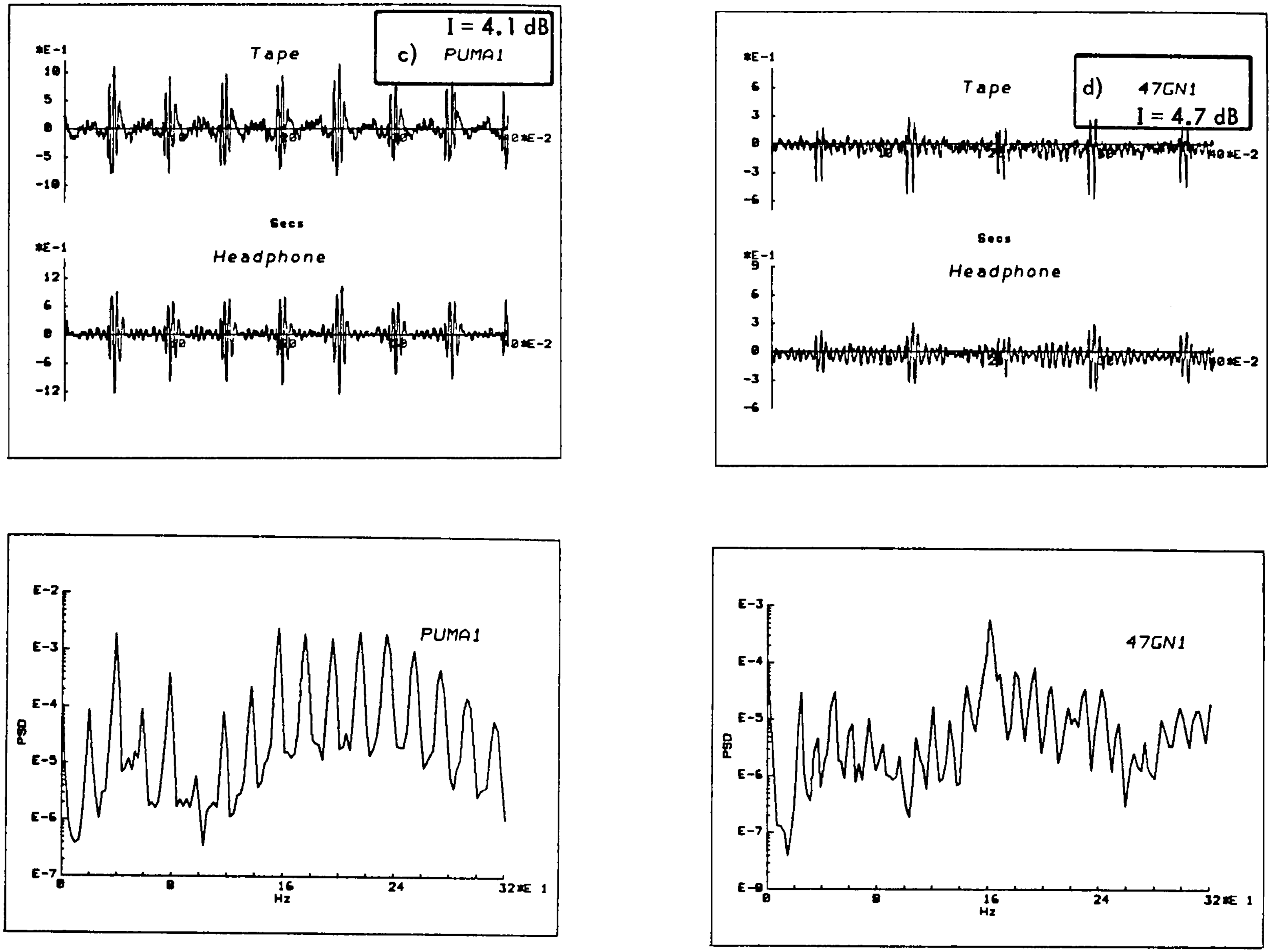


Figure C-2 (Continued)

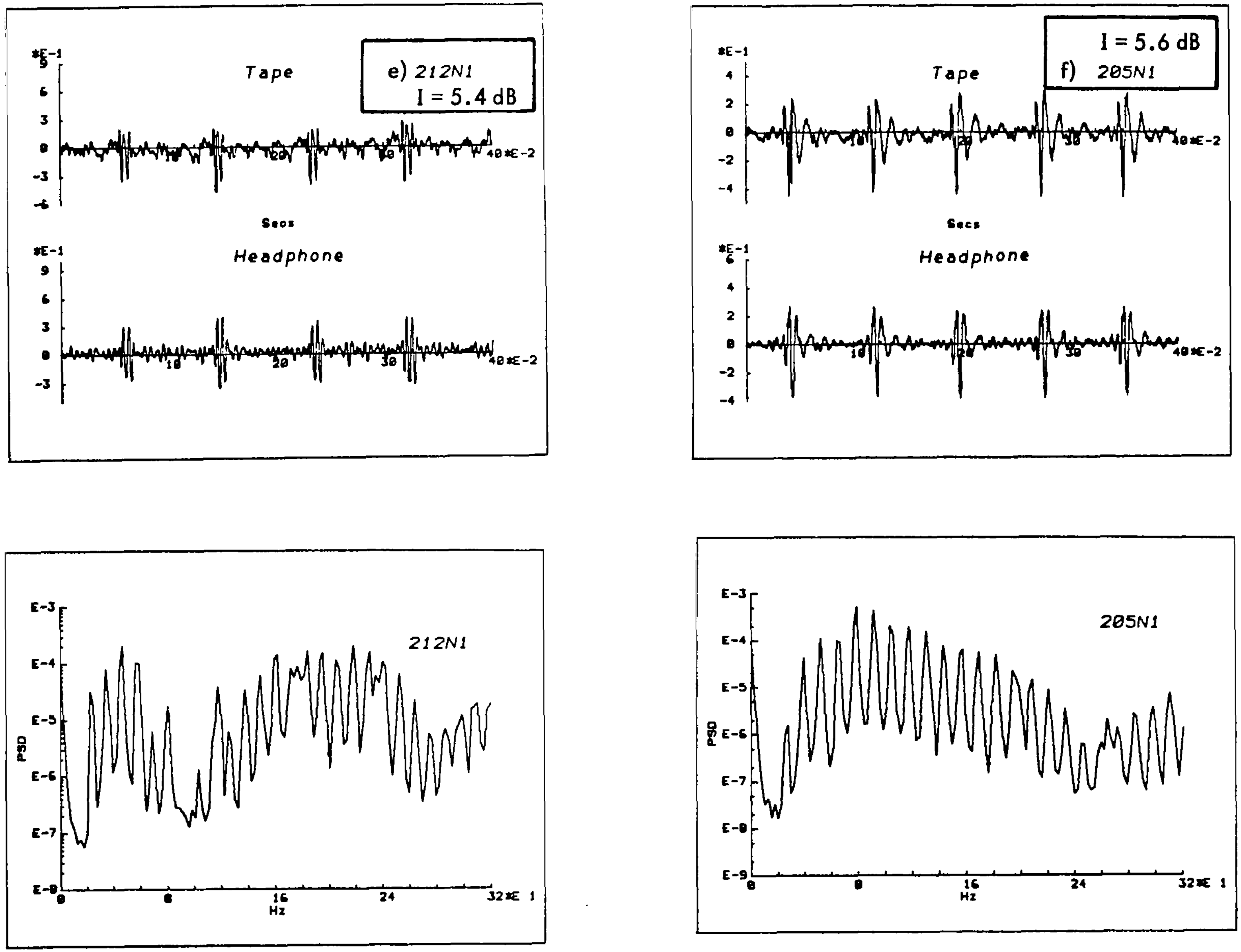


Figure C-2 (Continued)

Table C-1

Start Times and ISO-Impulse Corrections for Sample Sounds Plotted in Figures C-1 and C-2

Figure No.	Sound	Duration of Time History (sec)	Sample Start Time (sec)	$I = [EPNL_{ti} - EPNL_t]$, dB	
				Tape O/P	Coupler O/P
C-1a	SUFR 1	0.2	16	0.5	-
C-1b	350 N1	0.2	NA	0.6	-
C-1c	47G N2	0.2	7	0.8	-
C-1d	WSX 01	0.2	NA	1.7	1.7
C-1e	GAZ 02	0.2	5	2.7	-
C-1f	S64 N1	0.2	10	2.9	-
C-1g	S61 N1	0.2	16	3.7	3.2
C-1h	105 N2	0.2	7	3.9	2.6
C-1i	212 N3	0.2	NA	3.9	-
C-1j	C47 N3	0.2	7	4.6	1.8
C-2a	206 L2	0.4	17	2.2	-
C-2b	500 C1	0.4	10	3.7	1.7
C-2c	PUMA 1	0.4	5	4.1	3.6
C-2d	47G N1	0.4	12	4.7	-
C-2e	212 N1	0.4	15	5.4	4.4
C-2f	205 N1	0.4	15	5.6	3.2

APPENDIX D

Helicopter Characteristics

Table D-1

General Design Characteristics

Type No.	Model	Purpose	Max. Weight kg	Max. Speed km/hr	Engine No./Type	Engine kW
SA 321	Super Frelon	Multi	13,000	275	3 Turbine	3,470
SA 330	Puma	Transport	7,400	290	2 Turbine	2,350
SA 341/2	Gazelle	Utility	1,800	310	1 Turbine	440
SA 350	Ecurenil	General Purpose	1,900	272	1 Turbine	450
SA 365	Dauphin	General Purpose	3,000	315	1 Turbine	783
Messerschmitt-Boelkaw-Blohm Bo 105	-	Utility	2,300	270	2 Turbine	600
Westland	Sea King	ASW	6,000		2 Turbine	2,476
Westland	Lynx	General Purpose	2,600	280	2 Turbine	1,200
Bell 205	(UH-1H)	General Purpose	4,700	204	1 Turbine	820
Bell 206L	Long Ranger		1,800	241	1 Turbine	300
Bell 212	(UH-1N)	General Purpose	5,000	185	2 Turbine	960
Bell 47G	(2A and 5A)	General Purpose	1,340	170	1 Piston	165
Boeing CH47	Chinook	Transport	15,000	300	2 Turbine	5,600
Boeing CH53	-	Heavy Transport	19,000	300	2 Turbine	4,200
Sikorsky S61	Sea King	Transport	9,000	235	2 Turbine	2,240
Sikorsky S64	Tarke	Heavy Lift	19,000	203	2 Turbine	6,700
Hughes 500C	-	Utility	1,160	240	1 Turbine	300

Table D-2

Rotor Characteristics

Type No.	Model	Main Rotor			Tail Rotor		
		Speed rpm	Dia. (m)	B, No. of Blades	Speed rpm	Dia. (m)	B, No. of Blades
SA 321	Super Frelon	210	18.90	6	990	4.00	5
SA 330	Puma	265	15.0	4	1,278	3.04	5
SA 341/2	Gazelle	378	10.5	3	5,774	0.695	13
SA 350	Ecurenil		10.69	3		1.86	2
SA 365	Dauphin	348	11.50	4	4,700	0.90	13
Messerschmitt-Boelkaw-Blohm Bo 105	-		9.84	4		1.90	2
Westland	Sea King		18.90	5		3.16	6
Westland	Lynx		12.80	4		2.21	4
Bell 205	(UH-1H)		14.63	2		2.59	2
Bell 206L	Long Ranger	394	11.28	2	2,550	1.58	2
Bell 212	(UH-1N)	324	14.69	2	1,662	2.59	2
Bell 47G	(2A and 5A)	370	11.32	2	2,160	1.78	2
Boeing CH47	Chinook	245	18.29	3		--	-
Boeing CH53	-		22.02	6		4.88	4
Sikorsky S61	Sea King	203	18.90	5	1,136	3.23	5
Sikorsky S64	Tarke	186	21.95	6	852	4.88	4
Hughes 500C	-	484	8.03	4	3,110	1.30	2

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
NASA CR-3610		

4. Title and Subtitle	5. Report Date
LABORATORY STUDIES OF SCALES FOR MEASURING HELICOPTER NOISE	November 1982
	6. Performing Organization Code
7. Author(s)	8. Performing Organization Report No.
J. B. Ollerhead	WR 82-12
	10. Work Unit No.
9. Performing Organization Name and Address Wyle Research Wyle Laboratories 128 Maryland St. El Segundo, CA 90245	11. Contract or Grant No. NAS1-16276
	13. Type of Report and Period Covered Contractor Report
12. Sponsoring Agency Name and Address National Aeronautics and Space Administration Washington, DC 20546	14. Sponsoring Agency Code

15. Supplementary Notes

Langley technical monitor: Clemans A. Powell
Final Report

16. Abstract

A laboratory study has been made of the adequacy of the Effective Perceived Noise Level (EPNL) procedure for rating helicopter noise annoyance. Recordings of 89 helicopters and 30 fixed-wing aircraft (CTOL) flyover sounds were rated with respect to annoyance by groups of approximately 40 subjects. The average annoyance scores were transformed to Annoyance Levels defined as the equally annoying sound levels of a fixed reference sound. The sound levels of the test sounds were measured on various scales, with and without corrections for duration, tones, and impulsiveness. On average, the helicopter sounds were judged equally annoying to CTOL sounds when their duration-corrected levels are approximately 2 dB higher. Multiple regression analysis indicated that, provided the helicopter/CTOL difference of about 2 dB is taken into account, the particular linear combination of level, duration, and tone corrections inherent in EPNL is close to optimum. The results revealed no general requirement for special EPNL correction terms to penalize helicopter sounds which are particularly impulsive; apparently, impulsiveness causes spectral and temporal changes which themselves adequately amplify conventionally measured sound levels.

17. Key Words (Suggested by Author(s))		18. Distribution Statement
Helicopter Blade Slap	Helicopter Noise	Unclassified - Unlimited
Aircraft Noise	Impulsive Noises	
Psychoacoustics	Perceived Noise Level	Subject Category 71

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	126	A07

For sale by the National Technical Information Service, Springfield, Virginia 22161

NASA-Langley, 1982